



# **NAVAL POSTGRADUATE SCHOOL**

**MONTEREY, CALIFORNIA**

## **THESIS**

**COMPARISON OF DATA DEVELOPMENT TOOLS FOR  
POPULATING COGNITIVE MODELS IN SOCIAL  
SIMULATION**

by

Daniel C. McKaughan

September 2011

Thesis Advisor:

Christian Darken

Second Reader:

Jonathan Alt

**This thesis was performed at the MOVES Institute  
Approved for public release; distribution is unlimited**

THIS PAGE INTENTIONALLY LEFT BLANK

<b>REPORT DOCUMENTATION PAGE</b>			<i>Form Approved OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.				
<b>1. AGENCY USE ONLY (Leave blank)</b>		<b>2. REPORT DATE</b> September 2011	<b>3. REPORT TYPE AND DATES COVERED</b> Master's Thesis	
<b>4. TITLE AND SUBTITLE</b> Comparison of Data Development Tools for Populating Cognitive Models in Social Simulation			<b>5. FUNDING NUMBERS</b>	
<b>6. AUTHOR(S)</b> Daniel C. McKaughan				
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> Naval Postgraduate School Monterey, CA 93943-5000			<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> N/A			<b>10. SPONSORING/MONITORING AGENCY REPORT NUMBER</b>	
<b>11. SUPPLEMENTARY NOTES</b> The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number <u>NA</u> .				
<b>12a. DISTRIBUTION / AVAILABILITY STATEMENT</b> Approved for public release; distribution is unlimited			<b>12b. DISTRIBUTION CODE</b> A	
<b>13. ABSTRACT (maximum 200 words)</b> The United States is engaged in a new type of warfare. Defeating the enemy is now predicated on winning over local populations. To win these groups, commanders need to know what responses to expect for various operations in particular locations. Social simulations are a promising means of modeling these reactions, and there are several current methods used to populate these simulations with agents representative of a specific society. These methods, however, often require the input of subject matter experts and are costly in price and time. This thesis examines the simplification and automation of the agent instantiation process by conducting a usability study of two data development tools currently under consideration by the U.S. Army and TRAC-MTRY. The tools, a survey data case file generator developed at TRAC-MTRY and a text analysis tool (STANLEY) developed by Sandia National Laboratory, were examined in separate manners, and the results were encouraging. The survey tool was tested to validate in a practical manner its generated case files with respect to simulation output and real-world surveys. STANLEY was evaluated by scoring sentiment in a document corpus and attempting to correlate those scores to a real world issue. Results of the study indicate that the survey data tool generated case files of adequate quality to instantiate social simulations, potentially minimizing SME requirements and costs. Technical limitations precluded STANLEY from returning enough data for sufficient correlation comparison, although the results indicate the tool has potential.				
<b>14. SUBJECT TERMS</b> Social Simulations, Survey Data, Cultural Geography, Text Analysis, Agent Instantiation, Irregular Warfare, Data Development Tool			<b>15. NUMBER OF PAGES</b> 221	
			<b>16. PRICE CODE</b>	
<b>17. SECURITY CLASSIFICATION OF REPORT</b> Unclassified	<b>18. SECURITY CLASSIFICATION OF THIS PAGE</b> Unclassified	<b>19. SECURITY CLASSIFICATION OF ABSTRACT</b> Unclassified	<b>20. LIMITATION OF ABSTRACT</b> UU	

THIS PAGE INTENTIONALLY LEFT BLANK

**Approved for public release; distribution is unlimited**

**COMPARISON OF DATA DEVELOPMENT TOOLS FOR POPULATING  
COGNITIVE MODELS IN SOCIAL SIMULATION**

Daniel C. McKaughan  
Lieutenant Commander, United States Navy  
B.S., United States Naval Academy, 1996

Submitted in partial fulfillment of the  
requirements for the degree of

**MASTER OF SCIENCE IN  
MODELING, VIRTUAL ENVIRONMENTS, AND SIMULATION (MOVES)**

from the

**NAVAL POSTGRADUATE SCHOOL  
September 2011**

Author: Daniel C. McKaughan

Approved by: Christian Darken  
Thesis Advisor

Jonathan Alt, LTC, USA  
Second Reader

Mathias Kolsch  
Chair, MOVES Academic Committee

Peter Denning  
Chair, Computer Science Department

THIS PAGE INTENTIONALLY LEFT BLANK

## **ABSTRACT**

The United States is engaged in a new type of warfare. Defeating the enemy is now predicated on winning over local populations. To win these groups, commanders need to know what responses to expect for various operations in particular locations. Social simulations are a promising means of modeling these reactions, and there are several current methods used to populate these simulations with agents representative of a specific society. These methods, however, often require the input of subject matter experts and are costly in price and time. This thesis examines the simplification and automation of the agent instantiation process by conducting a usability study of two data development tools currently under consideration by the U.S. Army and TRAC-MTRY. The tools, a survey data case file generator developed at TRAC-MTRY and a text analysis tool (STANLEY) developed by Sandia National Laboratory, were examined in separate manners, and the results were encouraging. The survey tool was tested to validate in a practical manner its generated case files with respect to simulation output and real-world surveys. STANLEY was evaluated by scoring sentiment in a document corpus and attempting to correlate those scores to a real world issue. Results of the study indicate that the survey data tool generated case files of adequate quality to instantiate social simulations, potentially minimizing SME requirements and costs. Technical limitations precluded STANLEY from returning enough data for sufficient correlation comparison, although the results indicate the tool has potential.

THIS PAGE INTENTIONALLY LEFT BLANK

# TABLE OF CONTENTS

<b>I.</b>	<b>INTRODUCTION.....</b>	<b>1</b>
<b>A.</b>	<b>BACKGROUND .....</b>	<b>1</b>
	1. Irregular Warfare, Counterinsurgency Operations, and a New Center of Gravity .....	1
<b>B.</b>	<b>PROBLEM STATEMENT .....</b>	<b>3</b>
<b>C.</b>	<b>RESEARCH QUESTIONS.....</b>	<b>4</b>
<b>D.</b>	<b>APPROACH AND SCOPE.....</b>	<b>4</b>
<b>E.</b>	<b>CONTRIBUTIONS OF THE RESEARCH .....</b>	<b>4</b>
<b>F.</b>	<b>ORGANIZATION OF WORK.....</b>	<b>5</b>
<b>II.</b>	<b>SOCIAL SIMULATIONS AND TOOL OVERVIEW .....</b>	<b>7</b>
<b>A.</b>	<b>SOCIAL SIMULATIONS.....</b>	<b>7</b>
	1. Cognitive Theory.....	7
	a. <i>The Theory of Planned Behavior</i> .....	7
	b. <i>Fisher's Narrative Paradigm</i> .....	8
	2. Social Simulations Defined and Explained .....	8
	3. Feeding and Populating Social Simulations .....	9
	a. <i>Survey Data Development</i> .....	10
	b. <i>Current Methods of Survey Collection</i> .....	11
	c. <i>Leveraging a Document Corpus for Social Simulation</i> .....	11
<b>B.</b>	<b>THE CULTURAL GEOGRAPHY MODEL .....</b>	<b>13</b>
<b>C.</b>	<b>SURVEY DATA DEVELOPMENT TOOL .....</b>	<b>15</b>
	1. Issues with the Initial Version of the Tool .....	21
	2. The Tool Used in the Research .....	22
<b>D.</b>	<b>SANDIA'S TEXT ANALYSIS EXTENSIBLE LIBRARY (STANLEY).....</b>	<b>26</b>
	1. Constructing Classifiers .....	29
	2. Partitioning the Corpus.....	31
	3. Generating Case Files .....	31
<b>III.</b>	<b>METHODOLOGY .....</b>	<b>33</b>
<b>A.</b>	<b>CONSTRAINTS, LIMITATIONS, AND ASSUMPTIONS .....</b>	<b>33</b>
	1. Constraints.....	33
	2. Limitations.....	34
	3. Assumptions.....	35
<b>B.</b>	<b>SURVEY DATA DEVELOPMENT TOOL RESEARCH METHODS ...</b>	<b>36</b>
	1. Gathering the Survey Data .....	36
	2. Scrubbing and Cleaning the Data .....	38
	3. Establishing the Demographics and Organizing the Data .....	39
	a. <i>Age</i> .....	39
	b. <i>Community Type</i> .....	39
	c. <i>Education Level</i> .....	40
	d. <i>Income</i> .....	40

e.	<i>Political Party Affiliation</i> .....	40
f.	<i>Race</i> .....	40
4.	Adding Additional Responses to Account for Multiple Questions.....	41
5.	Establishing the Dynamic Inputs to the Surveys .....	42
6.	Culling the Surveys: Removing Insignificant Questions.....	46
7a.	Running the Unaltered Tool: Generating Case Files Using Modified Enumerated Responses .....	47
7b.	Running the Modified Tool: Generating Case Files Using Original Responses.....	49
8.	Creating the Bayesian Network shell .....	51
a.	<i>Adjusting the Survey Data to Meet Netica Restrictions</i> .....	51
b.	<i>Rerunning the Case Files with Corrections to Surveys</i> .....	52
9.	Scenario Development—Preparing the Case Files and the CG Input Tables for Use in the Model.....	55
a.	<i>Survey Tool Scenario</i> .....	55
10.	Methods of Analysis.....	56
C.	STANLEY RESEARCH METHODS .....	57
1.	Gathering Appropriate and Applicable Documents.....	59
a.	<i>Using a Web Spider, or Crawler, to Gather Documents</i> .....	61
2.	Developing the Classifiers .....	62
3.	Setting up and Executing STANLEY.....	63
4.	Establishing the Document Corpus .....	63
5.	Scoring the Documents .....	64
6.	Methods of Analysis.....	64
IV.	ANALYSIS .....	65
A.	SURVEY DATA DEVELOPMENT TOOL .....	65
B.	DISCUSSION OF ANALYSIS .....	75
C.	STANLEY.....	76
D.	DISCUSSION OF ANALYSIS .....	82
V.	CONCLUSIONS AND RECOMMENDATIONS.....	85
A.	SURVEY DATA DEVELOPMENT TOOL CONCLUSIONS AND RECOMMENDATIONS.....	86
B.	FUTURE WORK WITH THE SURVEY DATA DEVELOPMENT TOOL.....	86
C.	STANLEY TOOL CONCLUSIONS AND RECOMMENDATIONS .....	87
D.	FUTURE WORK WITH STANLEY.....	88
APPENDIX A.	SAMPLE PEW RESEARCH CENTER SURVEY .....	91
APPENDIX B.	PRELIMINARY WORK WITH STANLEY .....	121
A.	EARLY TESTS WITH STANLEY .....	121
B.	CASE STUDY .....	121
C.	RESULTS .....	122
D.	LESSONS LEARNED, RECOMMENDATIONS, AND FUTURE WORK .....	125

<b>APPENDIX C.</b>	<b>SEED DOCUMENTS FOR SPIDER RUNS .....</b>	<b>127</b>
<b>APPENDIX D.</b>	<b>EXAMPLE XML FILE GENERATED BY SPIDER .....</b>	<b>177</b>
<b>APPENDIX E.</b>	<b>ANALYSIS ON ORIGINAL STANLEY RESULTS.....</b>	<b>185</b>
<b>APPENDIX F.</b>	<b>SURVEY DATA DEVELOPMENT TOOL USERS GUIDE .....</b>	<b>191</b>
<b>LIST OF REFERENCES.....</b>		<b>197</b>
<b>INITIAL DISTRIBUTION LIST .....</b>		<b>201</b>

THIS PAGE INTENTIONALLY LEFT BLANK

## LIST OF FIGURES

Figure 1.	Contrasting conventional and irregular warfare (From DoD, 2007) .....	2
Figure 2.	Summary of data collection and scenario development methodology (After Baez, 2011) .....	9
Figure 3.	General approach to using a document corpus to populate models (From Baez, 2011) .....	12
Figure 4.	Document processing using a web crawler (From Baez, 2011) .....	13
Figure 5.	Graphical view of CG model (best viewed in color) (After Wiedemann, 2010) .....	15
Figure 6.	Survey Tool GUI (original version) showing selected target question and top four responses .....	17
Figure 7.	GUI for response delineation in survey tool .....	18
Figure 8.	Snapshot of command window while case files are being generated .....	19
Figure 9.	Improved Survey Data Tool Interface .....	23
Figure 10.	Stereotype Definition screen showing the inclusion of boxes to indicate a term as a stereotype .....	24
Figure 11.	Example case file showing stereotypes and other data .....	24
Figure 12.	Example ontology for population identities .....	29
Figure 13.	Example classification tree for partitioning a corpus. ....	30
Figure 14.	Research methodology for Survey Data Development tool .....	36
Figure 15.	Twenty year running track of satisfaction with current state of the nation (From Pew, 2010) .....	37
Figure 16.	Snapshot of results for Pew satisfaction question for 2010 .....	38
Figure 17.	Unemployment rate and economic confidence index for 2010 .....	43
Figure 18.	Unemployment rate and economic confidence index for 2010, reversed. Higher confidence index indicates worse consumer outlook. ....	44
Figure 19.	Oil Price data from 1997 through March 2011. Three major oil price indexes shown (best viewed in color) .....	44
Figure 20.	WTI Cushing (index used in thesis) Oil prices from 1997–March 2011 .....	45
Figure 21.	WTI Cushing Price Index for 2010 (\$) .....	45
Figure 22.	Sample JMP Pro 9 output used for culling survey questions .....	47
Figure 23.	Improved survey tool GUI. Includes ability to select stereotypes .....	50
Figure 24.	Netica generated Bayesian network illustration given demographics, coded indexes, and top three questions for entire survey .....	54
Figure 25.	Netica generated Bayesian network for Training Survey .....	54
Figure 26.	Research Methodology for Sandia’s text analysis tool .....	57
Figure 27.	WTI Cushing Price Index (From <a href="http://www.wtrg.com/daily/oilandgasspot.html">http://www.wtrg.com/daily/oilandgasspot.html</a> ) .....	58
Figure 28.	Early 2011 oil prices (WTI index) .....	59
Figure 29.	Brief summary of document collection process (After Baez, 2011) .....	60
Figure 30.	Generalized STANLEY process using spider .....	60
Figure 31.	Illustration of BN changes due to indicator shifts (best viewed in color) .....	66
Figure 32.	Twenty agent test results with reference lines showing index changes .....	68

Figure 33.	Relationship of most relevant demographics to satisfaction.....	69
Figure 34.	Satisfaction over time for CG2 for all replications (best viewed in color) .....	70
Figure 35.	Plot of satisfaction over time for each agent (CG2) (best viewed in color) ....	71
Figure 36.	Trained CG model satisfaction over time .....	72
Figure 37.	Trained CG model vs. Pew results over time (best viewed in color) .....	72
Figure 38.	Aggregated satisfaction over time for full model run (CG2 & CG3) (best viewed in color) .....	73
Figure 39.	Statistical analysis on significant differences between CG2 and CG3 .....	73
Figure 40.	Respondent/Agent Satisfaction percentages for CG1, CG3, and Pew Data sets (best viewed in color).....	74
Figure 41.	Aggregated satisfaction percentages (normalized) (best viewed in color) .....	75
Figure 42.	Oil prices and spider scores over time (normalized) (best viewed in color) ...	78
Figure 43.	Oil prices, forbes.com and all sites score (means, normalized) (best viewed in color) .....	78
Figure 44.	Mean and median scores with oil prices over time (normalized) (best viewed in color) .....	79
Figure 45.	Scatterplot matrix showing correlations of oil price to similarity scores .....	80
Figure 46.	Pair-wise correlations among oil prices and similarity scores.....	81
Figure 47.	Regression analysis for oil prices with mean similarity scores (Forbes and –All”).....	81
Figure 48.	Summary of fit results for oil prices and Forbes and –All” mean similarity scores (combined) .....	82
Figure 49.	Maximum similarity scores of websites over time (best viewed in color) .....	83
Figure 50.	Mean scores of websites over time (best viewed in color) .....	84
Figure 51.	Weka generated Bayesian Network showing the dependency of violence given genre, time written, and author gender .....	124
Figure 52.	Mean and median scores with oil prices over time (normalized) (best viewed in color) .....	185
Figure 53.	Oil prices, forbes.com and all sites score (means, normalized) (best viewed in color) .....	186
Figure 54.	Mean and median scores with oil prices over time (normalized) (best viewed in color) .....	186
Figure 55.	Scatterplot matrix showing correlations of oil price to similarity scores .....	187
Figure 56.	Pair-wise correlations among oil prices and similarity scores.....	187
Figure 57.	Regression analysis for oil prices with mean similarity scores (Forbes and All) .....	188
Figure 58.	Summary of fit results for oil prices and Forbes and All mean similarity scores (combined) .....	189
Figure 59.	Survey tool with ability to order terms .....	193
Figure 60.	Bin delineation GUI.....	195

## LIST OF TABLES

Table 1.	Summary of steps used to set up and execute survey data tool (original) .....	20
Table 2.	Improved Tool execution steps .....	25
Table 3.	Sample output of comparison of documents.....	27
Table 4.	Computing document similarity with cosine similarity function .....	28
Table 5.	Final demographic bins of respondents for survey tool.....	41
Table 6.	Demographics with enumerated values to the right.....	48
Table 7.	STANLEY results for number of documents and percentages for author gender and time period written. ....	123

THIS PAGE INTENTIONALLY LEFT BLANK

## LIST OF ACRONYMS AND ABBREVIATIONS

ABM	Agent-Based Model
ABS	Agent-Based Simulation
ASCOPE	Areas, Structures, Capabilities, Organizations, People, Events
BBN	Bayesian Belief Network
BN	Belief Network
BP	British Petroleum
BVI	Beliefs, Values, and Interests
CAS	Complex Adaptive System
CG	Cultural Geography (model)
COIN	Counterinsurgency
DoD	Department of Defense
GUI	Graphical User Interface
IW	Irregular Warfare
NGO	Non-Governmental Organization
PMESII-PT	Political, Military, Economic, Social, Information, Infrastructure, Physical Terrain, Time
SME	Subject Matter Expert
STANLEY	Sandia's Text ANaLysis Extensible librarY
SSTR	Security, Stability, Transition, and Reconstruction
SWEAT-MUS	Sewage, Water, Electricity, Academics, Trash, Medical, Unemployment, and Security
TPB	Theory of Planned Behavior
TRAC-MTRY	Training and Doctrine Command Analysis Center-Monterey
URL	Uniform Resource Locator
WTI	West Texas Intermediate
WTRG	West Texas Research Group
WVS	World Value Survey

THIS PAGE INTENTIONALLY LEFT BLANK

## **ACKNOWLEDGMENTS**

There are so many people to thank and acknowledge for their help with this thesis. First and foremost I need to thank my wife, Stacie, and my kids, Matthew and Anna. If it were not for their patience and understanding during all those periods of time when I was not home early or they saw me with my computer in my lap and knew not to bug me, I never would have gotten this finished. Thank you.

Thanks go to my brothers and sisters in Christ at Seaside First Baptist Church. Their prayers and encouragement have been a God-send during this time for me and my family

I must also thank my advisor, Dr. Chris Darken, and my second reader, LTC Jon Alt. Their guidance, advice, and patience was truly appreciated and an inspiration. Additionally, I would like to thank MAJ Francisco Baez, who tirelessly helped me with the Survey Data Tool and the Cultural Geography model. Those cheesecakes were well worth it.

Many thanks go to my friends and classmates that I have made while at NPS. The experience would not have been the same without any of you. In particular, I would like to thank MAJ Jason Caldwell, MAJ Mike Stinchfield, and Maj Don Herod, who were my cohort and my sounding board.

Finally, in no particular order, I would like to thank JT McClain and Zac Heath at Sandia National Laboratories for all their information, assistance, and insight into STANLEY; the Rolands and Associates Contractors at TRAC\_MTRY, Harold Yamauchi, Greg Sullivan and John Ruck, who helped me with fixing the many issues with the survey tool and in the implementation of the CG model; and my many professors and teachers here at NPS.

THIS PAGE INTENTIONALLY LEFT BLANK

# **I. INTRODUCTION**

## **A. BACKGROUND**

The last decade has seen a significant shift in the challenges faced by U.S. military forces and her allies. The types of battles being fought are not the same as those fought by previous generations. No longer is the battlefield a large space where soldiers, armor, and artillery engage in the mutual goal of annihilating each other. The skies overhead are no longer filled with dogfights at Mach 2. The recent wars in Iraq and Afghanistan have resulted in a dramatic shift in the way the United States approaches combat and military operations. In fact, a new type of warfare has developed. Irregular Warfare (IW) is defined as the “violent struggle among state and nonstate actors for legitimacy and influence over relevant populations” (DoD, 2007, p. 2).

### **1. Irregular Warfare, Counterinsurgency Operations, and a New Center of Gravity**

People, not platforms and advanced technology, will be the key to...success (DoD, 2007, p. 1).

The U.S. military is not shying away from conventional warfare. Quite the contrary, it understands that wars nowadays are multifaceted and multiregional. Wars and conflicts can take on any variation of intensity, longevity, and scope. Like conventional warfare, the goal of IW remains the same: win. Gain control of the forces, population, and territory of the enemy. While conventional warfare is state versus state and employs direct military confrontation to destroy an adversary, IW focuses on the influence of the local populations. IW is a political struggle with components that are both violent and nonviolent. The struggle, however, is not for outright destruction of enemy forces and its military machine, or for conquering of territories. This new struggle is for the influence and support of the relevant, local populations. The two sides in the conflict are both battling for the physical, psychological, and ideological support of the central, local population (DoD, 2007). Figure 1 shows the ideological shift from conventional to IW.

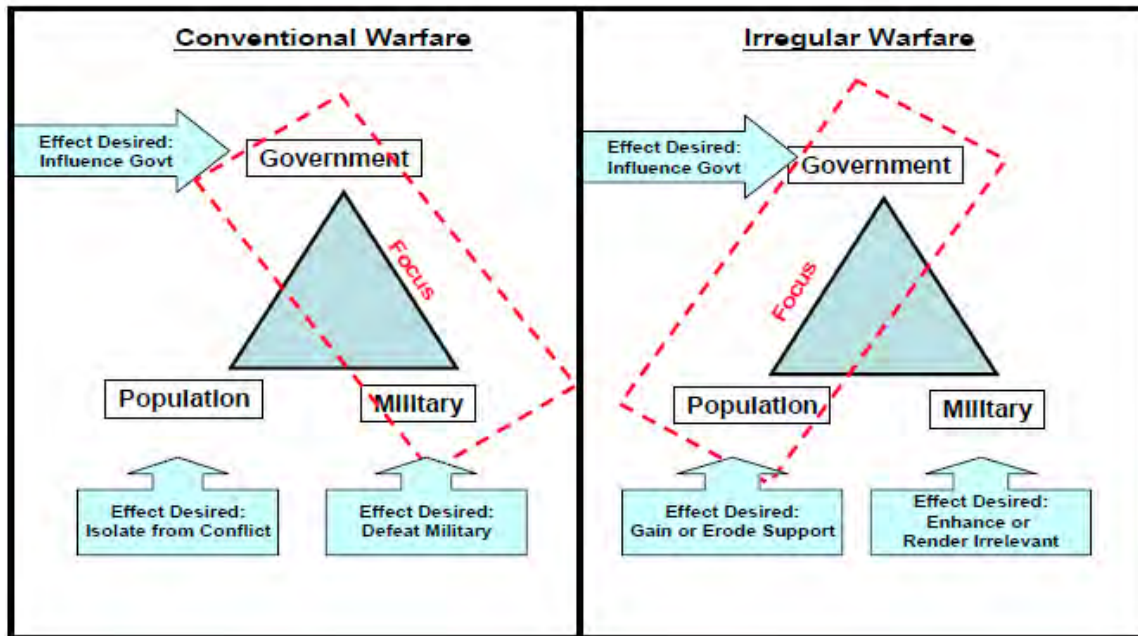


Figure 1. Contrasting conventional and irregular warfare (From DoD, 2007)

At the strategic level, IW is concerned with the security, stability, transition, and reconstruction (SSTR, now known as stability operations) of a region. Tactically, the focus is on counterinsurgency operations (COIN). As with any war, there has to be some goal, some target or area that should be the focus of all the combat effort. This is the center of gravity of the enemy. Clausewitz stated that the center of gravity is “the point against which all the energies should be directed” (Osgood, 2011). Mansoor and Ulrich (2007) build upon the centrality of the population when they conclude that the population is the new center of gravity for COIN and IW operations. They state that people are the key to success because people are the environment. Since the goals of insurgents and counterinsurgents are to separate the other group from the population while at the same time drawing the population’s support (whether actively or passively) themselves, whichever group wins the people wins the fight. From a United States and ally perspective, the success of IW and COIN rests, then, with the safeguarding of the population. Protect them from the kinetic effects of military operations, maximize the positive impacts of nonkinetic operations, and ultimately win over their support and trust (Alt, Jackson, Hudak, & Lieberman, 2009). The true challenge, though, and the vein in

which this thesis is written, is how to best go about winning the support and trust of the local populations. With so many operators vying for the local population's support, how does one group win out over another?

The answer is not at all a simple one. U.S. forces involved in the conflict need to have an understanding of the local population. This understanding needs to go beyond the simple idea of basic needs. The ability to determine how a society will respond to an event requires training and experience. Forecasting sentiment—trying to determine how a group of people will react to something—is exceptionally difficult, even in peaceful situations. How much more so in a conflicted environment, when the local populations are already at high mistrust levels?

In order to know how a society or group will react to an operation, and more importantly, how to get desired results in the population's response, the society has to be broken down into and understood from the most basic levels. Analysts need to be intimately familiar with a society's beliefs, attitudes, intentions, and behaviors (Alt, Everton, & Lieberman, 2010). One method to accomplish this is through creating social simulations that model societies. While exceptionally difficult to make perfect, these simulations can be extremely useful in gaining insight into a society's beliefs, values, and interests (BVI) and can lead to forecasts on possible futures of societal responses to events and operations.

## **B. PROBLEM STATEMENT**

The problem, as it applies to this research, is how to populate social simulations in an efficient and effective manner using disparate sources of information. How can models be instantiated in a manner that gives commanders and decision-makers adequate representations of social dynamics in insurgency warfare in a timely fashion while still providing analysts powerful methods to develop forecasted BVIs of a society? This thesis will discuss and analyze two of the tools used to generate societal belief networks that can be used to populate social simulations

## **C. RESEARCH QUESTIONS**

The research issues and questions that this thesis attempts to explain and answer are:

1. Can automation tools effectively assist in the generation and population of agents in cognitive models and social simulations?
  - a. Does the Survey Data Development Tool (or Survey Data Case File Generator) developed in-house at TRAC-MTRY simplify and enhance the creation of stereotyped case files for use in social simulations?
    - i. Do the resulting case files make sense in terms of the surveys examined and the scenario developed and tested?
    - ii. Does the output from the Cultural Geography (CG) model scenarios instantiated with the tool generated case files similarly compare to the results of the Pew research?
  - b. Can the text analysis tool STANLEY effectively aid in the generation of agent belief networks through sentiment analysis?
    - i. Can STANLEY correctly indicate a correlation between the narratives pertaining to some real world event and the corresponding effects on some other event?
2. How do model outputs compare to empirical data from selected real-world case studies?

## **D. APPROACH AND SCOPE**

The research presented here examines two independent social simulation population tools. The tools take as input some manner of societal beliefs (in this case, survey data and Internet articles that represent narratives) and produce case files and sentiment similarity scores that are used to instantiate the social simulations. This thesis is a usability study of the two tools, which are currently proofs of principal and parts of on-going research. The results of model runs using the tools' input will be compared to real world scenarios to check the practical validity of the tools as effective population methods for social simulations.

## **E. CONTRIBUTIONS OF THE RESEARCH**

Commanders in the field need information on possible operational effects on local populations as quickly as possible. Organizations like the U.S. Army's Training and Doctrine Command (TRADOC) Analysis Center (TRAC) provide that information. This

thesis will contribute significantly to the processes of populating social simulations in general and the Cultural Geography model specifically. Not only will it determine the usability of two tools currently under consideration by the U.S. Army and TRAC-MTRY and prove the viability of semi-automating the case file development and agent instantiation processes, it will also attempt to prove the viability of the tools' output as input for social simulation instantiation by comparing the output of the tools to real world results. Additionally, through semi-automation, organizations can expect to save money and resources by removing much of the extensive amount of subject matter expert (SME) involvement in the processes examined in this thesis.

## **F. ORGANIZATION OF WORK**

- Chapter I: Introduction. This chapter introduces the problem and the research questions to be answered. It lays out the plan of attack for the research and summarizes the organization of the paper.
- Chapter II: Social Simulation and Tool Overview. Here, the concepts of social simulations, cognitive theory, and model population are discussed. Additionally, the tools examined in this thesis are explicitly discussed.
- Chapter III: Methodology. This chapter defines the constraints, limitations, and assumptions of the work and describes the step-by-step processes undertaken in the setting up and execution of the research.
- Chapter IV: Analysis. The results of the model runs are analyzed and compared to real world data. Graphical representations of the data are explained and discussed.
- Chapter V: Conclusion and Recommendations. The final chapter summarizes the research and the results returned and makes recommendations regarding the benefits of the tools and possible areas of continued research.
- Appendix A: Sample Pew Research Center Survey. This is included so the user has an idea of how Pew conducts its surveys and the type of questions asked to respondents.
- Appendix B: Preliminary Work with STANLEY. Prior examination and experimentation with the STANLEY tool was conducted by the author and other research team members. The results of that study are included here and provide insight into the tool's behaviors.
- Appendix C: Seed Documents for the Spider. The documents used as STANLEY's classifier documents are included here.

- Appendix D: Sample XML File Generated by Spider. Included for the reader's curiosity, a sample xml file containing a website's text returned from the spider.
- Appendix E: Analysis of Original STANLEY Results. The tool returned similarity scores that included zero (0). These scores were excluded for the analysis conducted for the research. This appendix contains the analysis of the results including the scores of zero. This is here to show the difference in the two analyses.
- Appendix F: Survey Data Development Tool Users Guide. Part of the usability study of the survey data tool involved writing a user guide for the tool. This appendix includes that guide.

## II. SOCIAL SIMULATIONS AND TOOL OVERVIEW

As previously mentioned, this thesis will analyze and discuss two of the tools used to populate societal belief networks that can be used to populate social simulations. However, before that can be accomplished, the idea of social simulations needs to be defined and explained. Additionally, the social simulation used in this thesis, the Cultural Geography (CG) model, needs to be introduced, as do the two tools that are the focus of this research.

### A. SOCIAL SIMULATIONS

#### 1. Cognitive Theory

Attempting to explain human behavior and all its complexities is no easy task. Scientists and psychologists have been trying for centuries to understand how people make decision and why, and the approaches to this understanding are manifold (Ajzen, 1991). In order to simulate decision making and BVIs, however, this understanding must take place. Two methods of comprehending human behavior that are used in the development of social simulations are the Theory of Planned Behavior (TPB) and Fisher's Narrative Paradigm.

##### *a. The Theory of Planned Behavior*

TPB describes a manner in which people carry out actions based on intentions developed through beliefs. Specifically,

TPB accesses an individual's 1) belief toward a particular behavior, 2) belief about the social norms associated with a particular behavior, and 3) belief regarding the ability to control the outcome of a particular behavior. These are referred to as behavioral beliefs, normative beliefs, and control beliefs, respectively, and together yield the individual's level of intention to carry out a particular action. (Alt & Lieberman, 2010b, p. 127)

This intention that the individual develops is thought to be the precursor to actual action that he or she might take. If the intentions can be modeled, then the actions

can be predicted. This is where social simulations can be beneficial. The actions of a society as a whole can be predicted by determining the sums of the intentions of the individuals in the group.

***b. Fisher's Narrative Paradigm***

Fisher's paradigm can serve as a foundational concept for the populating of social simulations and models. Fisher (1989) believes that each person possesses a unique identity based on his/her culture and experiences, and that this identity is presented through the narratives and stories he/she tells. These stories, essentially, are the individual's view of the world in which he/she lives and of the other people he/she encounters in it. The collection of all these narratives, then, can be representative of the population of which the individuals are part.

**2. Social Simulations Defined and Explained**

Social simulations, as a whole, attempt to represent large human populations as complex adaptive systems (CAS). A CAS is a complex, self-similar collection of interacting adaptive agents ("Complex Adaptive Simulations," 2010). These representations can be at varying levels of detail and granularity, but the overarching goal of action/response prediction remains the same. Regardless of the type of social simulation, the key goals remain the representation of BVIs of individuals in a population and the changes that occur in their BVIs across a range of events (Alt, Jackson, Hudak, & Lieberman, 2010).

The make-up of a social simulation consists of actors which are representations of the individuals comprising the strata of the population. In order to accurately develop the social simulation, each entity must be informed of the issues relevant within its group, know the rules for interaction within its society, and be able to formulate intention in order to carry out actions (Alt et al., 2010). The simulation must also contain a representation of the social environment within which the actors "live" and "interact" (National Research Council, 2008).

A benefit of social simulations and simulated societies is that they provide tools for analysts to gain insights into the complexities of populations. As previously mentioned, social simulations are made up of entities, or agents. These agents perceive and influence aspects and areas of their environments (including other agents), and they may learn. Emerging from the complex interactions among the agents, patterns may develop (a society is born), and these interactions can provide explanations for social phenomena (Sun, 2007). Ultimately this is the goal: To use social simulations to explain social phenomena, and from that explanation, predict societal responses to events in order to plan and conduct COIN and IW operations with the bi-fold results of attaining national strategic and tactical security objectives while improving local conditions and opinions.

### 3. Feeding and Populating Social Simulations

Figure 2 illustrates an example of the process used to gather data and develop a scenario for use in a model such as the Cultural Geography model. This is a broad reaching example, as it lists many of the methods used to collect data, but both polling data and narratives are mentioned and considered important methods for gathering regional data on a population for use in modeling.

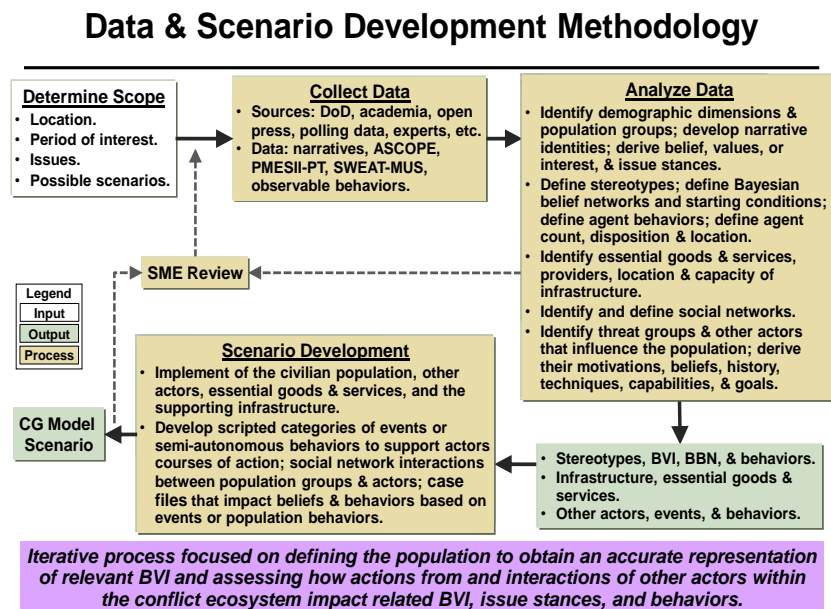


Figure 2. Summary of data collection and scenario development methodology (After Baez, 2011)

In order for the social simulation to work, in order for it to be of any use in providing insight and feedback about a society's beliefs and reactions, it must be populated, or fed, with useful and reliable information from which it can derive authentic representations of human behavior and develop cognitive agents. There are numerous ways of instantiating the models, but this thesis will focus on two. The first utilizes quantitative survey and polling data of local populations to generate Bayesian belief networks (BBN) that can be used to populate the models. The second uses the idea of the narrative paradigm, that societies are made up of agents whose BVIs are expressed through their stories and what they have to say, to develop narrative ethnographies (Alt & Lieberman, 2010b).

#### *a. Survey Data Development*

The use of polls and surveys to predict outcomes and sentiment is nothing new. These methods have been around for centuries to aid researchers, analysts, politicians, etc. in drawing conclusions and making decisions regarding how society feels about a specific issue or how it will behave or respond to a certain event. Using survey and polling data in the population of social simulations requires flexibility in the methods used to garner the data. Since there are no survey instruments executed on a recurring basis used solely for the purpose of social simulation (Alt & Lieberman, 2010b), analysts and model developers must look elsewhere to find the data they need. This is an ongoing effort across multiple disciplines. The tools and techniques to leverage the less common polling and survey data are well known (Kalton, 1983; Kiecolt & Nathan, 1985; Alt & Lieberman, 2010; Lieberman & Alt, 2010), though work is required to collect the data and tailor it for use with social simulations.

There are some strengths and limitations to survey use. It is extremely difficult and untenable to directly observe a population's behavior over the time scales of interest. However, the models can be informed by sample observations and self reporting data. While sample observations are certainly possible for very specific behaviors in concise contexts (e.g., employees using the stairs or the elevators at the office), self report (social surveys) tends to be easier to conduct. There are of course limits to self report. It

is prone to errors as respondents may have memory lapse or misunderstanding of the questions. Furthermore, respondents can directly deceive the questionnaire, providing false or misleading answers. This effect can be mitigated with insurances of anonymity and confidentiality, but variance in error rates is still difficult to determine (Alt & Lieberman, 2010b). Finally, the collection of valid polling data within conflict environments will always be somewhat problematic (McKaughan, Alt, Heath, & McClain, 2011).

#### ***b. Current Methods of Survey Collection***

The tools and techniques used to collect survey data are widely known and well established. The World Value Survey (WVS) is a long term social and behavioral research project that is striving to determine and describe the cross-cultural values spanning 62 countries. The detailed questionnaires contain over 250 items and are mostly associated with socio-cultural, political, religious, and moral views of the respondents (“World Values,” 2010). The WVS has been conducted in waves since 1981.

Other popular social survey organizations include Gallup, which has been “studying human nature and behavior for more than 75 years” ([www.gallup.com/corporate/115/About-Gallup.aspx](http://www.gallup.com/corporate/115/About-Gallup.aspx), 2010), the European Social Survey ([www.europeansocialsurvey.org](http://www.europeansocialsurvey.org)), the Pew Research Center ([www.people-press.org/](http://www.people-press.org/)), and the United States General Social Survey ([www.norc.org/GSS+Website/](http://www.norc.umd.edu/GSS+Website/))

#### ***c. Leveraging a Document Corpus for Social Simulation***

The other method of representing a society that this thesis will look at is through the use of a document corpus, or a body of documents (texts, websites, media, blogs). Using the principles of the narrative paradigm, a collection of narratives (documents) is assumed to be representative of the agent’s BVIs in a simulation. By collecting large amounts of literature and writings from a population and comparing the documents to a set of test documents which have been predetermined to be representative of a society, it is possible to construct BNs for use in social models.

Like survey data, this too has some weaknesses and limitations. While the number of documents could be quite large, retrieving and sorting the corpus can be quite time consuming. Determining and developing the comparison test sets often requires subject matter expertise. The cognitive model framework has traditionally been manually populated. This manual process requires the assistance of a SME in a particular field to build the model of that field. This can be a very time consuming process (Bauer et al., 2005). Also, it is still possible to deceive and bias the outcome (Basilico et al., 2008). The media can be quite effective at changing the opinions of a group, so media related documents should not be the primary input to a model.

The overall general approach to using a collection of narratives, or a document corpus, is illustrated in Figure 3. Simply stated, the documents are collected by some method (a web crawler, for example). Then, they are categorized using methods described in II.C and Figure 4. These categorized documents form a training set of documents that are used as a comparison for other documents collected. Belief networks are established based on the results of the comparison, and the agents in the model are then initialized.

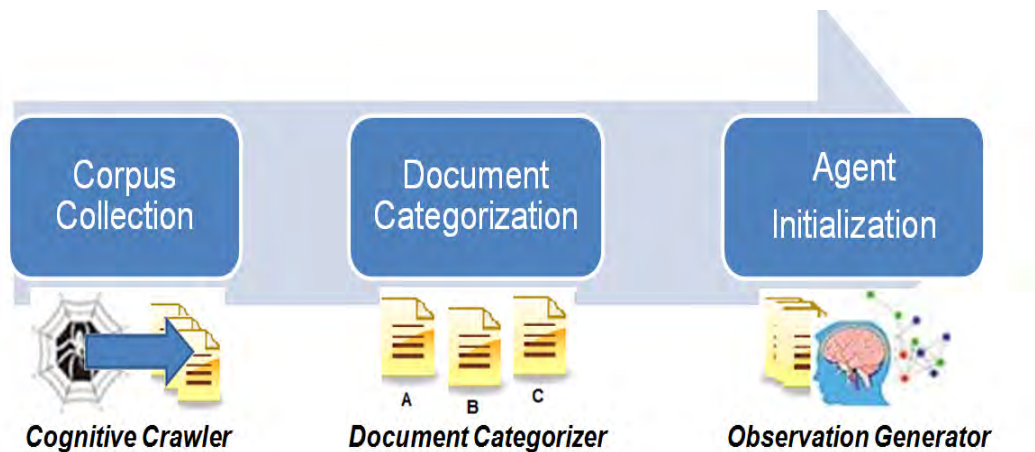


Figure 3. General approach to using a document corpus to populate models (From Baez, 2011)

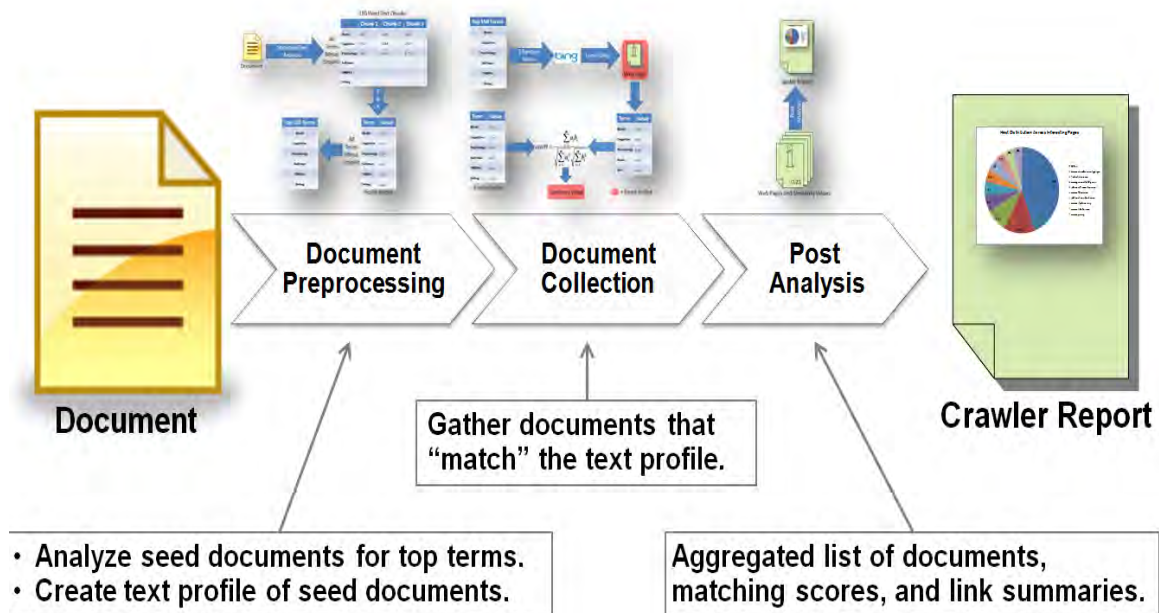


Figure 4. Document processing using a web crawler (From Baez, 2011)

## B. THE CULTURAL GEOGRAPHY MODEL

The CG model is a government owned, open source agent-based model (ABM) of the operational environment based on doctrine and the interpretation of social networks. It is a population-centric model developed to support the analysis of societal behaviors focusing on population perception based upon friendly and threat actions. The purpose of the model is to address the responses of a population in a conflict environment (Wiedemann, 2010). Developed at TRAC-MTRY, the CG model is designed to provide insight into the potential effects U.S. operations will have on the BVIs of local civilian populations. It provides the following benefits:

- Provides the commander with information toward developing an understanding of how the population responds to actions within the area of operations.
- Supports course of action (COA) development by characterizing potential futures of population stances on multiple lines of effort.

- Data and scenario development process provides framework toward a thorough understanding of indigenous population.
- Versatile “interest driven” modeling architecture capable of receiving many forms of data about the population.

Patterned after the conflict eco-system described by Kilcullen (2007) in an attempt to capture the complexities of irregular warfare, the model uses various theories of social science to allow decision makers to track changes to positions at various sizes of the population (Alt & Lieberman, 2010). Figure 5 summarizes the CG model.

The CG model consists of agents or entities that represent people in a group or population. These “people” are modeled to interact with each other and respond to specific events. Every agent is defined by a set of demographic dimensions that collectively shape the entity’s BVIs and stances on issues such as security, governance, etc., and behaviors. “The model captures the behavioral response of the civilian population in the environment of conflict, which can provide valuable insights regarding actions or events under consideration by the coalition forces” (Wiedemann, 2010). Fisher’s narrative paradigm (Fisher, 1987), described above, is underlying social theory upon which narrative identities are developed to form agent BVIs (Hudak et al., 2010).

The key aspect of the CG model is that it allows users and analysts to schedule within the model events that have the potential to impact the beliefs and stances of populations of interest through the use of Bayesian networks. These scheduled events could range from the murder of a community leader by insurgents to a Non-Governmental Organization (NGO) constructing a building or providing a service (Hudak et al., 2010). Additionally, the CG model supports representation of a social network by:

Applying concepts of propinquity (physical proximity) and homophily (tendency to associate with those of similar interests). Specifically, when an entity’s belief changes, the entity attempts to communicate the result to other entities within a pre-defined physical distance that possess similar interests. If communication is successful, the receiver’s beliefs are impacted accordingly. (Hudak et al., 2010, p. xvi)

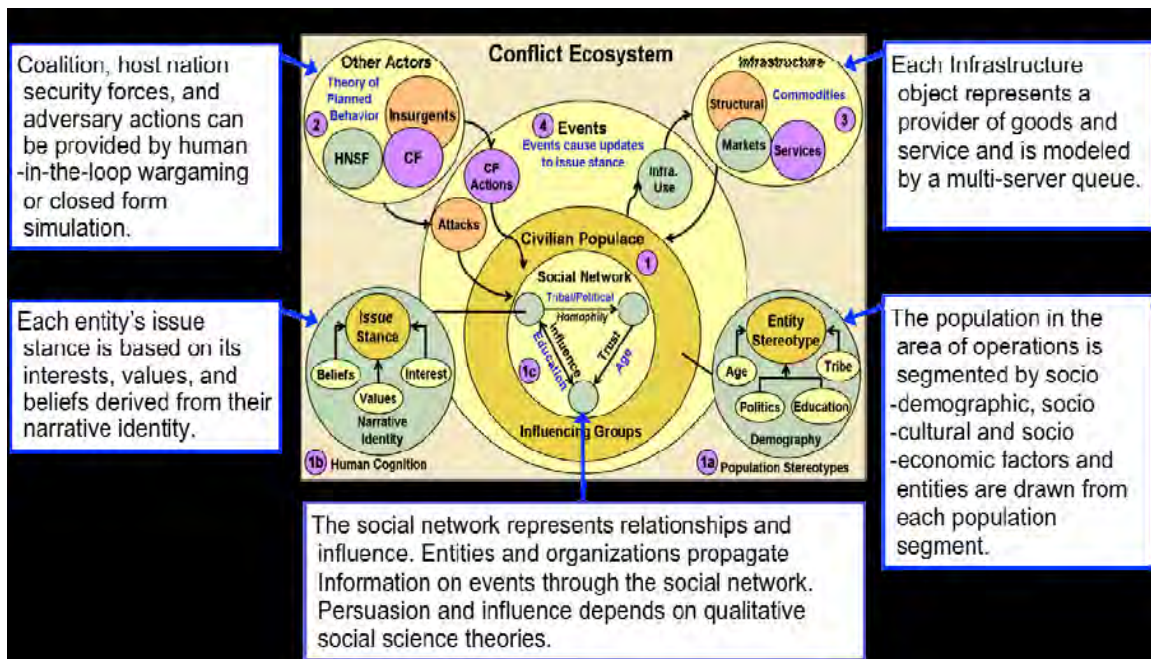


Figure 5. Graphical view of CG model (best viewed in color) (After Wiedemann, 2010)

### C. SURVEY DATA DEVELOPMENT TOOL

The survey data development tool, also known as the survey data case file generator, was also developed in house at TRAC-MTRY. The primary purpose of the tool is to generate case files. In their simplest forms, case files represent an agent's beliefs, implemented in spreadsheet form. Case files feed the underlying conditional probability tables within the issue networks, resulting in changes to end node probabilities (issue stances). They serve two main purposes: (1) establishing starting condition probabilities for an entity's issue network, and (2) changing the probabilities of an entity's issue network due to external events (such as a natural disaster or terrorist bombing) and population behaviors (such as acquiring essential services). Traditionally, case files were generated manually through SME intervention and assistance. SMEs completed surveys of questions about beliefs, issues, and values of a group of people and how those groups might react to certain events. This SME survey data helped inform case file development.

The Java-based development tool analyzes survey data and develops case files based on a score(s) related to a question or term's dependency(ies) on other terms or

questions within the survey and ranks according to which is most significant. These case files are then used to develop stereotypes of a population based on the survey results, and they represent the final step of data development for instantiating issue stances. One or more survey files can be selected for loading. Once loaded, the user has the option of choosing the type of data he is working with. The default is categorical, but this can be changed to continuous if needed. A target can also be selected. Setting a target is important, as this will be the term or question that is of interest. The targeted term will not be ranked in the final tool output since its values are used to calculate two-way Chi-squared classification for the other categorical terms. Continuous terms are ranked by variance.

If no target is selected, then the distributions for the categorical terms can be customized. The default values are evenly distributed based on the number of categorical terms. For example, if a term has values of either 1 or 2, then the tool expects 50% of the terms to be valued at 1 and 50% of the terms to be valued at 2. Figure 6 shows the survey tool graphical user interface (GUI), Figure 7 the GUI for delineating the bin values, and Figure 8 shows the command screen while case files are being generated.

Once the specification and customization are complete, the tool will read through all the loaded files and output a ranking of the terms. Again, the categorical terms are ranked based on Chi-square values and continuous terms by variance. By default, only the top four ranked terms are selected, but this is modifiable as the user may choose any or all of the terms. The tool will only generate case files for selected terms. The user may then customize the bins for each term selected. The tool by default assumes six responses per term and assigns values to those terms. These values should be modified to reflect the responses' values in the survey. If the question in the survey has more or less than six terms, the user may alter the number of responses and adjust the output values accordingly. For example, if the survey contains the demographic response of race, and gives the respondent four choices (white, black, Hispanic, other), the user has the option of decreasing the number of responses from six to four, appropriately naming each bin, and inputting minimum and maximum values for each response. A solid, easily recognizable naming convention is essential, as these names will form the stereotype

names. A naming convention for the current example would be R1, R2, R3, R4 (representing white, black, Hispanic, and other, respectively). After the case files are created, stereotypes of the population can be generated based on the results.

Note: the tool was modified extensively over the course of the thesis progress. The above description is for the original version of the tool. Section II.C.2 describes the changes made to the tool. Section III.B describes the process of tool utilization and case file generation that revealed the shortfalls and resulted in the changes made to the tool throughout the research.

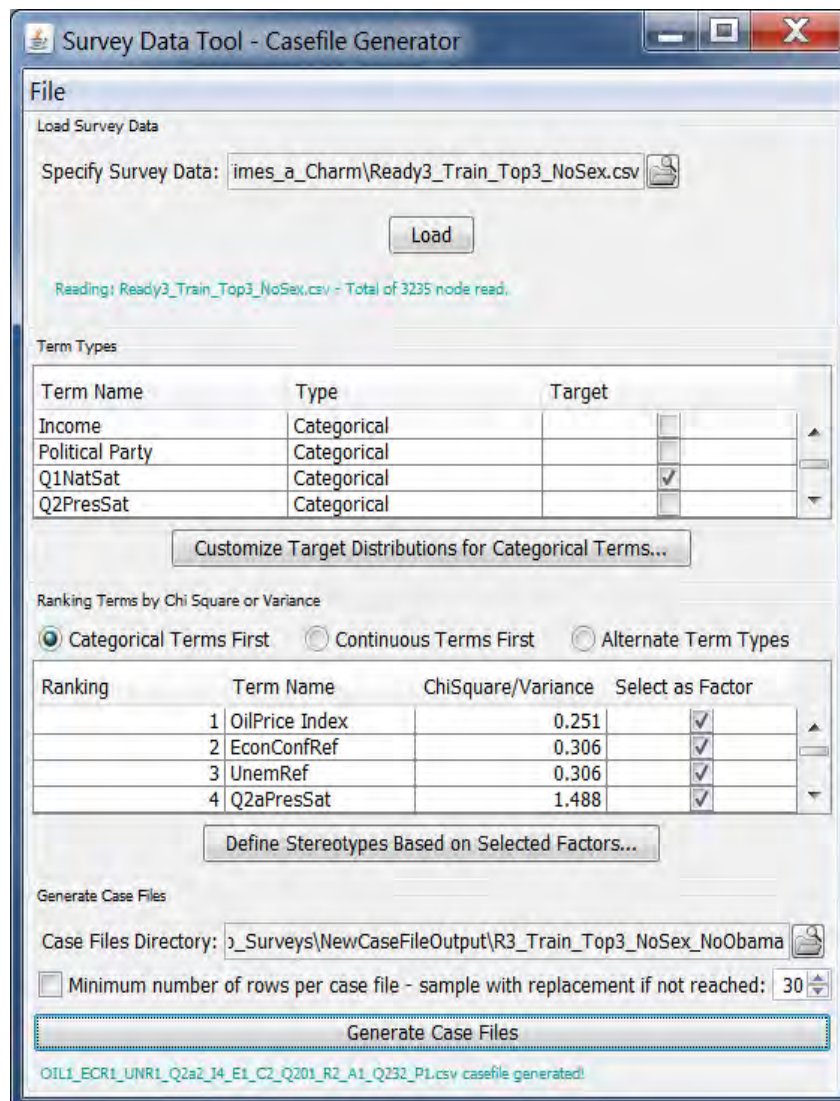


Figure 6. Survey Tool GUI (original version) showing selected target question and top four responses

Demographic Stereotype Definition

☐ Use each value to designate a demographic bin

Number of Demographic Bins: 2

Long Name	Short Name	Range (Low)	Range (High)
Q2aPresSalVery s...	Q2a1	Very strongly	Very strongly
Q2aPresSalNot so...	Q2a2	Not so strongly	Not so strongly

Income Category

☐ Use each value to designate a demographic bin

Number of Demographic Bins: 4

Long Name	Short Name	Range (Low)	Range (High)
Income75_150K	T3	75_150K	75_150K
IncomeLT40K	T1	LT40K	LT40K
IncomeGT_Eq150K	T4	GT_Eq150K	GT_Eq150K

Education Category

☐ Use each value to designate a demographic bin

Number of Demographic Bins: 3

Long Name	Short Name	Range (Low)	Range (High)
EducationHighSc...	E2	HighSchool	HighSchool
EducationGTCol...	E4	GTCollege	GTCollege
EducationLTHigh...	E1	LTHighSchool	LTHighSchool

Comm Type Category

☐ Use each value to designate a demographic bin

Number of Demographic Bins: 3

Long Name	Short Name	Range (Low)	Range (High)
CommTypeUrban	C1	Urban	Urban
CommTypeSubur...	C3	Suburban	Suburban
CommTypeRural	C2	Rural	Rural

Q20GovEffect Category

☐ Use each value to designate a demographic bin

Number of Demographic Bins: 3

Long Name	Short Name	Range (Low)	Range (High)
Q20GovEffectSum...	Q202	Sumw	Sumw
Q20GovEffectGood	Q203	Good	Good
Q20GovEffectNone	Q201	None	None

Race Category

☐ Use each value to designate a demographic bin

Number of Demographic Bins: 3

Long Name	Short Name	Range (Low)	Range (High)
RaceWhite	R1	White	White
RaceOther	R3	Other	Other
RaceBlack	R2	Black	Black

Age Category

☐ Use each value to designate a demographic bin

Number of Demographic Bins: 3

Long Name	Short Name	Range (Low)	Range (High)
AgeMiddle Age	A2	Middle Age	Middle Age
AgeSenior	A3	Senior	Senior
AgeYoung	A1	Young	Young

Q23TaxBurden Category

☐ Use each value to designate a demographic bin

Number of Demographic Bins: 3

Long Name	Short Name	Range (Low)	Range (High)
Q23TaxBurdenLess	Q231	Less	Less
Q23TaxBurdenMore	Q233	More	More
Q23TaxBurdenA...	Q232	AboutRight	AboutRight

Political Party Category

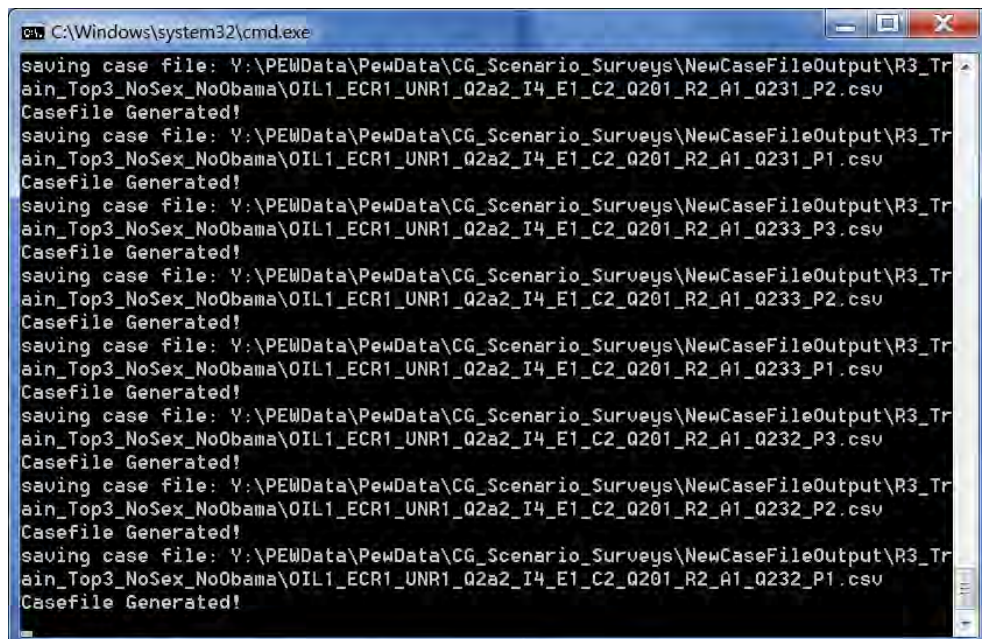
☐ Use each value to designate a demographic bin

Number of Demographic Bins: 3

Long Name	Short Name	Range (Low)	Range (High)
Political PartyInde...	P3	Independent	Independent
Political PartyRep...	P2	Republican	Republican
Political PartyDem...	P1	Democrat	Democrat

Apply Cancel

Figure 7. GUI for response delineation in survey tool



```
C:\Windows\system32\cmd.exe
saving case file: Y:\PEWData\PewData\CG_Scenario_Surveys\NewCaseFileOutput\R3_Train_Top3_NoSex_NoObama\OIL1_ECR1_UNR1_Q2a2_I4_E1_C2_Q201_R2_A1_Q231_P2.csv
Casefile Generated!
saving case file: Y:\PEWData\PewData\CG_Scenario_Surveys\NewCaseFileOutput\R3_Train_Top3_NoSex_NoObama\OIL1_ECR1_UNR1_Q2a2_I4_E1_C2_Q201_R2_A1_Q231_P1.csv
Casefile Generated!
saving case file: Y:\PEWData\PewData\CG_Scenario_Surveys\NewCaseFileOutput\R3_Train_Top3_NoSex_NoObama\OIL1_ECR1_UNR1_Q2a2_I4_E1_C2_Q201_R2_A1_Q233_P3.csv
Casefile Generated!
saving case file: Y:\PEWData\PewData\CG_Scenario_Surveys\NewCaseFileOutput\R3_Train_Top3_NoSex_NoObama\OIL1_ECR1_UNR1_Q2a2_I4_E1_C2_Q201_R2_A1_Q233_P2.csv
Casefile Generated!
saving case file: Y:\PEWData\PewData\CG_Scenario_Surveys\NewCaseFileOutput\R3_Train_Top3_NoSex_NoObama\OIL1_ECR1_UNR1_Q2a2_I4_E1_C2_Q201_R2_A1_Q233_P1.csv
Casefile Generated!
saving case file: Y:\PEWData\PewData\CG_Scenario_Surveys\NewCaseFileOutput\R3_Train_Top3_NoSex_NoObama\OIL1_ECR1_UNR1_Q2a2_I4_E1_C2_Q201_R2_A1_Q232_P3.csv
Casefile Generated!
saving case file: Y:\PEWData\PewData\CG_Scenario_Surveys\NewCaseFileOutput\R3_Train_Top3_NoSex_NoObama\OIL1_ECR1_UNR1_Q2a2_I4_E1_C2_Q201_R2_A1_Q232_P2.csv
Casefile Generated!
saving case file: Y:\PEWData\PewData\CG_Scenario_Surveys\NewCaseFileOutput\R3_Train_Top3_NoSex_NoObama\OIL1_ECR1_UNR1_Q2a2_I4_E1_C2_Q201_R2_A1_Q232_P1.csv
Casefile Generated!
```

Figure 8. Snapshot of command window while case files are being generated

Table 1. Summary of steps used to set up and execute survey data tool (original)

	Set Up Step	Description/Instructions	Image
1.	Open survey data tool	Execute via command window or .bat file in appropriate directory	
2.	Specify survey data	Click on search icon to locate survey to be used by the tool (this can be one survey or all in a given directory)	See Figure 6
3.	Select target or customize the target distributions	Click on the question that is the desired target, or set up the distributions for the questions	Figure 6
4.	Load the survey	Click the “Load” button  This will activate the tool and scored and ranked results will populate the “Ranking Terms by Chi Square or Variance” window	Figure 6
5.	Select factors upon which to build stereotypes	Click the boxes that contain the desired terms or questions that stereotypes should be built upon, or select all	Figure 6
6.	Ensure proper bin labeling	Click on the “Define Stereotypes Based on Selected Factors” button  Can choose default settings, or name and put proper responses in	Figures 6 and 7
7.	Load the stereotypes	Click the “Apply” button	Figure 7
8.	Determine where case files will be out put	Click the search icon to locate the directory and/or folder where case files will be deposited upon creation	Figure 6
9.	Set minimum number of rows	Check the box to set minimum number of rows in case file and set value	Figure 6
10.	Generate case files	Click on the “Generate Case Files” button	Figures 6 and 8

## **1. Issues with the Initial Version of the Tool**

Initially, the tool had some significant drawbacks that greatly affected its ease of use and its ability to handle surveys containing multiple response types. (Survey responses can be numbers, rankings, string responses, etc.) First, it did not take string data, only non-negative integers. This proved unacceptable as most survey data (and nearly all used here) have responses that are strings. Second, the tool did not take negative numbers. The dynamic index of consumer economic confidence (defined in III.B.5 below) is a negative number based index. Both of these drawbacks were discovered by trial and error, as there was no documentation for the tool. Finally, it was not possible to save individual tool application set up. For example, significant work goes in to delineating the responses for each question and ensuring the tool knows how many responses are available for each question and what those responses are. This effort, which must go in to each application of the tool, cannot be saved, and this greatly increases the amount of time doing work that has already been completed. While these drawbacks will be discussed in greater detail later in Chapter III, it is necessary to mention them here in order to understand the changes in survey preparation and processing that had to occur if the original version of the tool were to be used.

- Limitation: The tool would not accept string responses
  - Effect: All survey responses that were strings had to be given numerical equivalencies and the survey data had to be rewritten to account for and include numerical responses
- Limitation: The tool would not accept negative numbers
  - Effect: Certain dynamic indexes that are negative number based had to be rewritten to be positive number based. These new indexes had to be added to existing survey data
- Limitation: The tool would not save set up data and updated bin delineations

- Effect: If mistakes were made, or the tool needed to be rerun using previously used survey data, the user had to reenter all the bin names and values

The tool, in its original form, was deemed to be too time consuming and required too much alteration of the survey data in order to be effective and of any beneficial use. At the request of TRAC-MTRY, Gregory Sullivan, an analyst and computer software developer employed by Rolands and Associates, made alterations to the tool's code to rectify the issues discussed above. The tool will now take string responses, in addition to both positive and negative numerical values. Additionally, the tool will now save prior sessions, as long as the tested surveys' configuration files reside in the same directory as the execution file. These configuration files are automatically generated when the user executes the tool. Mr. Sullivan also made many additional tweaks to the tool's code in order to facilitate its use. These alterations will be discussed in the subsequent sections where they were noticed.

## **2. The Tool Used in the Research**

Given the issues discussed above, and the corrections made by Mr. Sullivan, an improved version of the tool had to be used for the generation of case files used in populating the CG model. Several other changes were made to streamline the tool and the case files generated. First, after the user selects the terms of interest to a particular target question, he can now select which of those terms are to form the stereotypes for case file generation and which terms solely represent the data that will be included in the case files. Figures 9 and 10 show the improved tool's interface and bin delineation screens. There is a new option to select all the terms in the ranking and a "Stereotype" box can be checked to mark that term as a stereotype. If a term is checked as a stereotype, it will be used in case file generation and will be a part of the case file's name. Remaining terms not designated as stereotypes will indicate which data will be included in the stereotype. Figure 11 shows a sample case file that includes data not checked as a stereotype. The case file columns are generated in order of the ranked terms, with the

target as the first column. Section III.B.3 describes the stereotype demographics used in this thesis while III.B.5–7 describes the other data included in the case files.

Table 2 includes the steps to operate the tool with the additional improvements included. Appendix D. contains an up-to-date Survey Data Development Tool User Guide. It is highly recommended to reference the User Guide for the particulars regarding saving the tool’s operating settings and using these saved files.

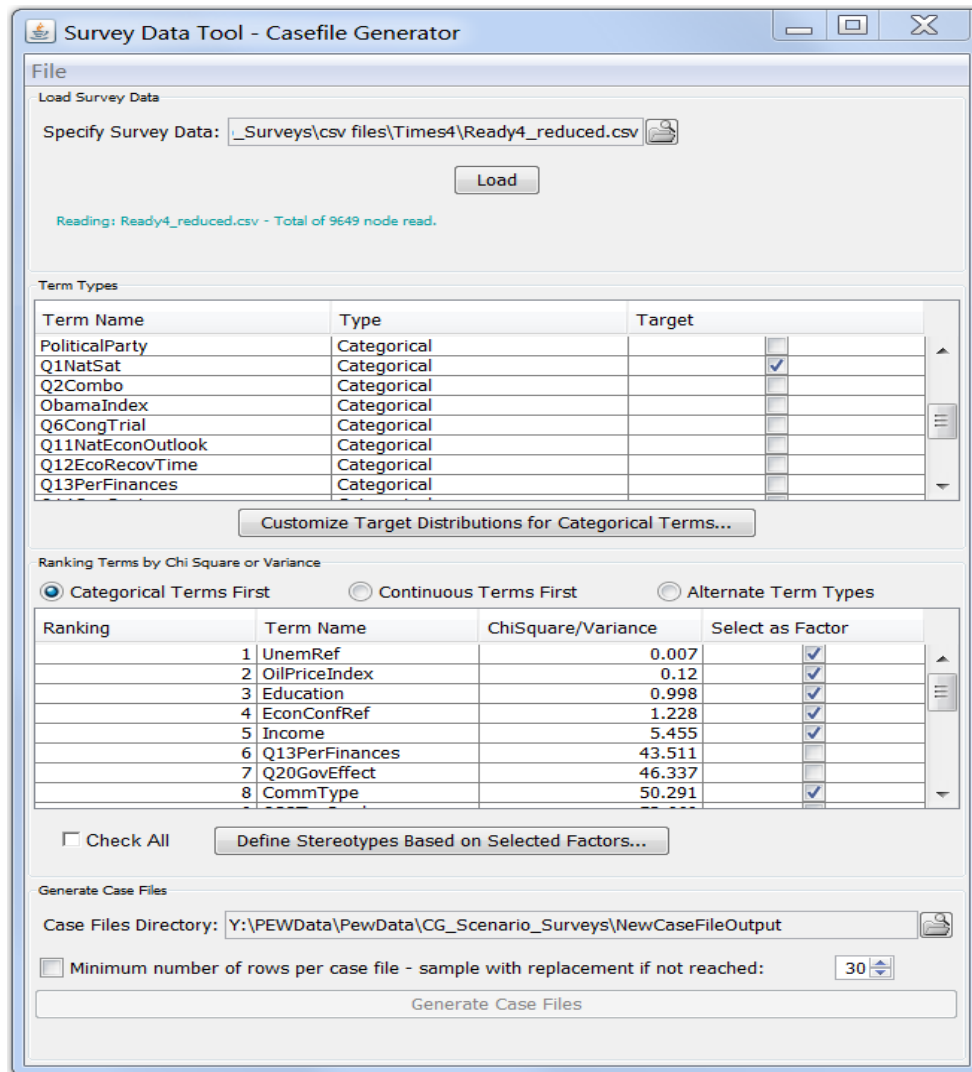


Figure 9. Improved Survey Data Tool Interface

**Demographic Stereotype Definition**

**OilPriceIndex Category**

☒ Use each value to designate a demographic bin ☐ Stereotype

Number of Demographic Bins: 6

Long Name	Short Name	Range (Low)	Range (High)
OilPriceIndex...	OINegTrend	NegTrend	NegTrend
OilPriceIndex	OIPosTrend	PosTrend	PosTrend

**Education Category**

☐ Use each value to designate a demographic bin ☒ Stereotype

Number of Demographic Bins: 4

Long Name	Short Name	Range (Low)	Range (High)
EducationLT...	E4	GI College	GI College
EducationLT...	E1	LTHighSchool	LTHighSchool

**EconConfRef Category**

☒ Use each value to designate a demographic bin ☐ Stereotype

Number of Demographic Bins: 6

Long Name	Short Name	Range (Low)	Range (High)
EconConfRef...	ECNegTrend	NegTrend	NegTrend
EconConfRef	ECPosTrend	PosTrend	PosTrend

**Income Category**

☐ Use each value to designate a demographic bin ☒ Stereotype

Number of Demographic Bins: 4

Long Name	Short Name	Range (Low)	Range (High)
IncomeI1	I1	I1	I1
IncomeI4	I4	I4	I4

**Q13PerFinances Category**

☒ Use each value to designate a demographic bin ☐ Stereotype

Number of Demographic Bins: 6

Long Name	Short Name	Range (Low)	Range (High)
Q13PerFinan...	Q1		
Q13PerFinan	Q1Fair	Fair	Fair

**Q20GovEffect Category**

☒ Use each value to designate a demographic bin ☐ Stereotype

Figure 10. Stereotype Definition screen showing the inclusion of boxes to indicate a term as a stereotype

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Q1NatSat	UnemRef	OilPriceIndex	Education	EconConfRef	Income	Q13PerFinances	Q20GovEffect	CommType	Q23TaxBurden	Age	Race	PoliticalParty
2	Dissatisfied	NegTrend	NegTrend	LTHighSchool	NegTrend	I1			Urban		Senior	White	Democrat
3	Dissatisfied	PosTrend	NegTrend	LTHighSchool	PosTrend	I1			Urban		Senior	White	Democrat
4	Dissatisfied	NegTrend	NegTrend	LTHighSchool	NegTrend	I1		Some	Urban	More	Senior	White	Democrat
5	Dissatisfied	NegTrend	NegTrend	LTHighSchool	NegTrend	I1		Some	Urban	AboutRight	Senior	White	Democrat
6	Dissatisfied	NegTrend	PosTrend	LTHighSchool	PosTrend	I1		Some	Urban	More	Senior	White	Democrat
7	Dissatisfied	PosTrend	PosTrend	LTHighSchool	PosTrend	I1			Urban		Senior	White	Democrat
8	Satisfied	PosTrend	NegTrend	LTHighSchool	PosTrend	I1	Fair		Urban		Senior	White	Democrat
9	Satisfied	PosTrend	PosTrend	LTHighSchool	PosTrend	I1	Good		Urban		Senior	White	Democrat

Figure 11. Example case file showing stereotypes and other data

Table 2. Improved Tool execution steps

	Set Up Step	Description/Instructions	Image
1.	Open Survey data tool	Execute via command window or .bat file in appropriate directory	
2.	Specify survey data	Click on search icon to locate survey to be used by the tool (this can be one survey or all in a given directory)	See Figure 9
3.	Select target or customize the target distributions	Click on the question that is the desired target, or set up the distributions for the questions	Figure 9
4.	Load the survey	Click the “Load” button  This will activate the tool and scored and ranked results will populate the “Ranking Terms by Chi Square or Variance” window	Figure 9
5.	Select factors upon which to build stereotypes	Click the boxes that contain the desired terms or questions that stereotypes should be built upon, or select all	Figure 9
6.	Ensure proper bin labeling	Click on the “Define Stereotypes Based on Selected Factors” button  Can choose default settings, or name and put proper responses in	Figures 9 and 10
7.	Choose stereotype factors	Deselect “Use each value” button and check “Stereotype” button  Note: data from terms not selected as stereotypes will appear in case files	Figure 10
8.	Load the stereotypes	Click the “Apply” button	Figures 7 and 10
9.	Determine location where case files will be saved	Click the search icon to locate the directory and/or folder where case files will be deposited upon creation	Figure 9.
10.	Set minimum number of rows	Check the box to set minimum number of rows in case file and set value, if desired	Figure 9
11.	Generate case files	Click on the “Generate Case Files” button	Figures 8 and 9

#### **D. SANDIA'S TEXT ANALYSIS EXTENSIBLE LIBRARY (STANLEY)**

As previously discussed, it is a possibility that the sentiment of a population or society is represented by the collected body of narratives that the particular society of interest generates. As introductory groundwork to this thesis, McKaughan, Alt, Heath & McClain (2011) took an extensive preliminary look at STANLEY in an attempt to determine its viability as a tool to populate social simulations through the generation of Bayesian networks.

Sandia's Text Analysis Extensible Library (STANLEY) is a text analysis library developed under Sandia's Cognitive Science & Technology program. STANLEY builds on decades of research in the fields of information retrieval, linguistics, and machine learning to analyze the statistics of a concept (in this case a body of narrative documents) used by an individual or group and thereby discover implicit semantic relationships among those concepts. Concepts can be representative of nearly anything that can be extracted from text, including words, phrases, geospatial locations, and abstract notions (e.g., hostile intent).

STANLEY supports a wide range of document types as input from which it can extract text to build a profile. Each profile consists of a searchable repository of documents along with statistical information about each document and the group of documents as a whole. STANLEY employs a bag-of-words approach to create a vector-space model (Salton, Wong, & Yang, 1975). The vector-space model represents each document as a vector of concept activations. The relative weight of each concept in the vector is calculated using a modified version of the log-entropy method (Equation 1.1) (Dumais, 1991; Ribeiro-neto & Baeza-Yates, 1999), which assigns higher values to concepts that are better at distinguishing between documents (Salton & Buckley, 1991).

STANLEY is quite versatile in its ability to handle varieties of text. It eliminates words that do not convey much information such as words with less than four characters, words with more than 18 characters, words that contain a combination of letters and numbers, and commonly used stop-words. Furthermore, STANLEY is capable of handling text in non-English languages or other alphabets, as long as a list of stop-words

can be provided. The produced textual profiles can be used to measure the similarity of individual or groups of documents by comparing their concept vectors using a cosine similarity measurement. Tables 3 and 4 demonstrate the comparison of two documents using cosine similarity. This process begins by converting both documents into their associated term vectors. These two vectors are then compared using the cosine similarity function shown in Equation 1.3 (Berry et al., 2003; Ko & Berry, 2004; Bauer, Laham, Benz, Dooley, & Kimmel, 2005).

Table 3. Sample output of comparison of documents

	Doc 1	Doc 2	Doc 3
<b>Model</b>	0.57	0.76	
<b>Cognitive</b>	0.12	0.84	
<b>Psychology</b>	0.32		
<b>Software</b>			0.23
<b>UMBRA</b>		0.91	0.43
<b>Debug</b>			0.75

$$\text{Log-Entropy} = \log(1 + f_y) \left( 1 + \left( \frac{\sum_j (p_y \log(p_y))}{\log n} \right) \right) \quad (0.1)$$

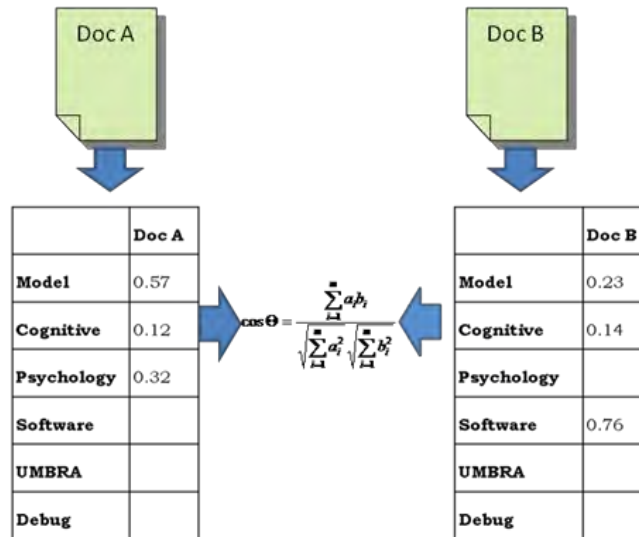
Where  $f_{ij}$  is the frequency of term  $i$  in document  $j$  and  $p_{ij}$  is:

$$p_{ij} = \frac{f_{ij}}{g_i} \quad (0.2)$$

and  $g_i$  is the number of document in which  $i$  occurs.

$$\cos \Theta = \frac{\sum_{i=1}^m a_i b_i}{\sqrt{\sum_{i=1}^m a_i^2} \sqrt{\sum_{i=1}^m b_i^2}} \quad (0.3)$$

Table 4. Computing document similarity with cosine similarity function



One of the most important features of STANLEY is the ability to create cognitive models from textual profiles. In order to do this, STANLEY makes the assumption that each concept (word, phrase, etc.) in a textual profile corresponds to a concept in a cognitive model. In other words, words or phrases in one document can be found in another. The idea of co-occurrence is then used to populate the network. Co-occurrence describes how often two words or phrases occur within the two documents. The co-occurrence can be in paragraphs, chunks or whole documents (Basilico et al., 2008).

Social simulations need to represent the beliefs and values of a society or group of people. Bayesian belief networks are one method used to represent those aspects in a social simulation. As a matter of fact, mathematicians and philosophers have been using probability theory to describe human cognition for well over 200 years (Alt, Jackson, Hudak, & Lieberman, 2009). To utilize the STANLEY tool in the development of Bayesian Belief networks, some preparation work has to be done to the document corpus, in particular in the development and establishment of the test body, otherwise known as classifiers.

## 1. Constructing Classifiers

The test body of documents with which the document corpus for the group under analysis will be compared needs to be divided into a series of classifiers. The process to partition and accurately analyze the corpus relies on the construction of classifiers which represent and describe the categories for each partition of the data set. The categories represent branches in a tree that divides the document corpus into appropriate bins based on comparisons to a classifier. The specific categories used in each scenario might vary based on the geographic area of interest and the issues to be modeled. Figure 12 is an example tree of factored categories.

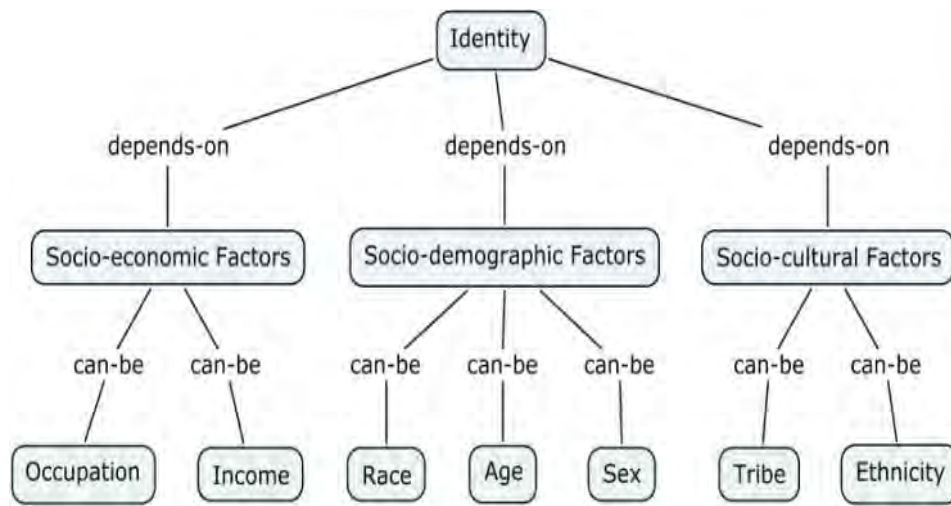


Figure 12. Example ontology for population identities

Further detail is often required to refine and define categories within the dimensions chosen from the terminal nodes of Figure 12 for representation within the scenario. Partitioning of the corpus could take a sequential approach, given appropriate documents as classifiers. In order to accurately detail the group or society in question, it may be required to further define categories, as shown by Figure 13.

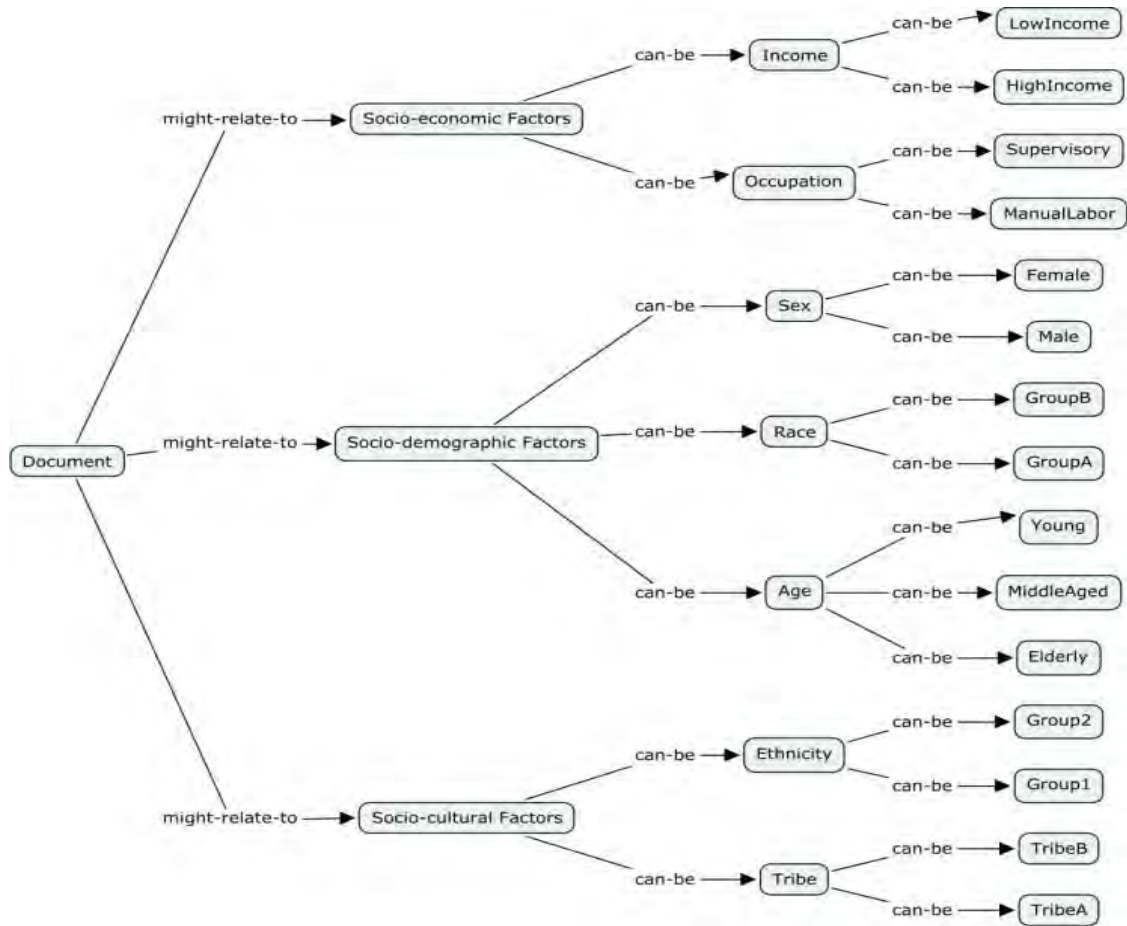


Figure 13. Example classification tree for partitioning a corpus.

The classifiers are constructed by providing example documents for each category value within the hierarchy. STANLEY is used to build a textual profile for the category values based on these example documents. Corpus documents will be compared to the documents in the classifier list and then given a category value classification score based on how well they match the category value profile. This score is taken directly from the cosine similarity comparison of the document's concept vector with the concept vector of the category value's corpus. A profile and subsequent scoring is also generated for each category as a whole. The matching scores of a document and a set of thresholds are used to determine the category observations for that document and used later in the belief network generation process. The level of confidence of the subsequent partitioning steps is directly related to the level of confidence in the documents used to construct the classifiers.

## **2. Partitioning the Corpus**

The partitions chosen for the corpus depend on the needs of the scenario and should be geared to inform cognitive models for the relevant population subtypes. Socio-demographic, socio-economic, and socio-cultural factors are typically used to describe the population. These same factors can be used to partition the corpus. The other overlying theme in social simulation is the concept of salient issues, or the determination of what issues are of significant importance to a population (Alt, Jackson, Hudak, & Lieberman, 2009). Ideally, the cognitive models would be developed that represent the society's stance on these defined issues.

If the document is related to socio-cultural factors, for example, it can be related to ethnicity or tribe. The document would then be checked for similarity with one of the terminal bins in this example, being given a strength of association with tribe A or tribe B, for instance. Once a document has been associated with the terminal bins in the tree in Figure 12, the set of documents associated with a particular composite entity can be classified by relevant issue and stance. For example, the documents related to young, male, low-income, manual laborers from racial group A, tribe B and ethnic group 2 can be classified by their level of agreement with a categorizer related to satisfaction on security or a categorizer related to dissatisfaction with security. Through this sequential process the analyst provides a quantitative level of confidence regarding the appropriateness of the documents used to inform each representative entity stereotypes stance on a given issue.

## **3. Generating Case Files**

In order to create the belief networks, each document is treated as an observation of the represented population. These document observations are created using the category matching scores generated by the classifiers. Every document is given a score for each category and category value. The scores are passed through a simple, configurable threshold function, also known as the configuration file, to assign a category value to each category. The threshold functions set the minimum scores that a document must possess in a given category. These minimum scores are arbitrarily established by

the user. It allows checks to see if the document matches the profile for the category as a whole with a given confidence. It then checks each category value score and categorizes the document based on the largest. The user can also establish a score delta (difference between the highest score and the values of all the other classifier scores for that particular classifier) that must be exceeded in order for the category to be matched. Any document that does not meet the threshold criteria will be given an assignment of UNKNOWN (represented as a “?” in the output file) for that category. After the full categorization stage, there is a list of document observations and each document provides an observed category value for each category. These observations allow one to observe correlations of category values across the population. The observations can be fed to a machine learning tool such as Weka to determine the conditional probabilities of a pre-defined Bayesian belief network or learn the structure of a Bayesian belief network directly. This type of machine learning is ideal for the case when there is missing observation data, such as when a category value of UNKNOWN is assigned. The generated belief networks can now be used to initialize the agent-based model. It needs noting that the quality of the generated Belief networks is highly dependent on both the appropriateness of the classifiers and how well the gathered corpus represents the population being modeled.

Appendix B describes the results of some preliminary experimentation with STANLEY.

### **III. METHODOLOGY**

As mentioned, this thesis looked at two different tools that could be of potential benefit in the instantiation and application of the CG model. The overall methodology for this thesis was relatively straight forward. At the outset, two separate scenarios were developed and examined which would test the potential usefulness and benefits of each of the two model populating tools. First, data was collected for each tool. Survey and polling data concerning the country's overall satisfaction with the state of the nation were gleaned from the Pew Research Center for the People and the Press for use in the survey data development tool. Historical oil price data as well as documents relating to the early 2011 turmoil and unrest in the Middle East and North Africa were gathered to run and test the STANLEY tool. Once the data for each tool was gathered, collated, and cleaned, scenarios were developed for use in the CG model to test the survey tool. The CG model output generated with the tool's case files was examined and compared to the known results of the surveys to test the practical validity of the tool. For STANLEY, the document scores generated through tool application were compared to the oil prices to determine the existence of a correlation. Each tool's set up and execution is described in detail in this chapter.

#### **A. CONSTRAINTS, LIMITATIONS, AND ASSUMPTIONS**

As with any usability study or research project, there are certain constraints, limitations, and assumptions that the researcher sets and/or must work within to accomplish the goals of the study. This section describes those associated with each of the tools analyzed. Additionally, it discusses the limitations of the data sets used in the research.

##### **1. Constraints**

- STANLEY
  - STANLEY was run offsite, at Sandia's lab in Albuquerque, NM; this prevented direct interaction with the tool and

oversight of the operation. Additionally, any changes to be affected to the tool over the course of the research could not be made.

- Average time for a spider crawl was three to five days per website, with an additional one to two days for similarity computations and analysis, resulting in over a week per website processing.
- CG Model
  - In order to control the behavior of the model to deterministic behavior, distributions utilized within the model's input tables were held constant.

## **2. Limitations**

- Survey Data Tool
  - The tool is a prototype, with research ongoing. As such, development and upgrades occurred concurrent with this research and were limited to those issues discovered during its use that could be brought to the attention of and corrected by contractors assigned to TRAC-MTRY.
  - No case studies of the tool existed prior to this research. There were no previous experiences with the tool to use as guidelines for shaping and conducting this study.
- STANLEY
  - Like the survey tool, STANLEY is on-going research. While other studies have been conducted in the area of sentiment forecasting and semantic analysis, it is still a proof of concept, particularly in the area of social simulations and model population.
- Data Limitations
  - The survey data used in this research was concatenated from numerous, varying surveys conducted by the Pew Research Center.
  - As such, not all respondents had answers to all the questions (not all questions were asked in every survey). This resulted in answers with no responses (blanks) in the surveys used for this research.
  - The Pew data contained no sampling error to compare results.

- CG Model
  - Run time was a significant limiting factor. Number of agents and replications had to be kept relatively low to minimize model run time.
  - Runs took two days to over ten, depending on number of replications.

### 3. Assumptions

- STANLEY returned similarity scores of zero (0). After consulting with Sandia, the zero scores are present for one of two main reasons. First, there were no words in common between the two documents (except for common stop words). Second, the tool could not score the document because it was an image or some other text free object. The former is extremely unlikely, so this thesis assumes the latter for scores of zero (0).
- CG Model: Numerous assumptions were made for the CG model scenarios used with the survey tool.
  - Constant distributions of actions in the model reflected population attitudes and were reflected in satisfaction shifts.
  - Only three indicators affected the agents' satisfaction. These were Oil Prices, Unemployment Rate, and Economic Confidence Index (all described in III.5).
  - The shifts in satisfaction created by changes in these indexes were enough to reflect the actual Pew survey results.
  - 1,296 stereotypes (generated by the survey tool) were enough to populate the CG model.

## B. SURVEY DATA DEVELOPMENT TOOL RESEARCH METHODS

This section lays out the research methodology that was used in analyzing and evaluating the survey data development tool.

### Methodology-Survey Data Tool

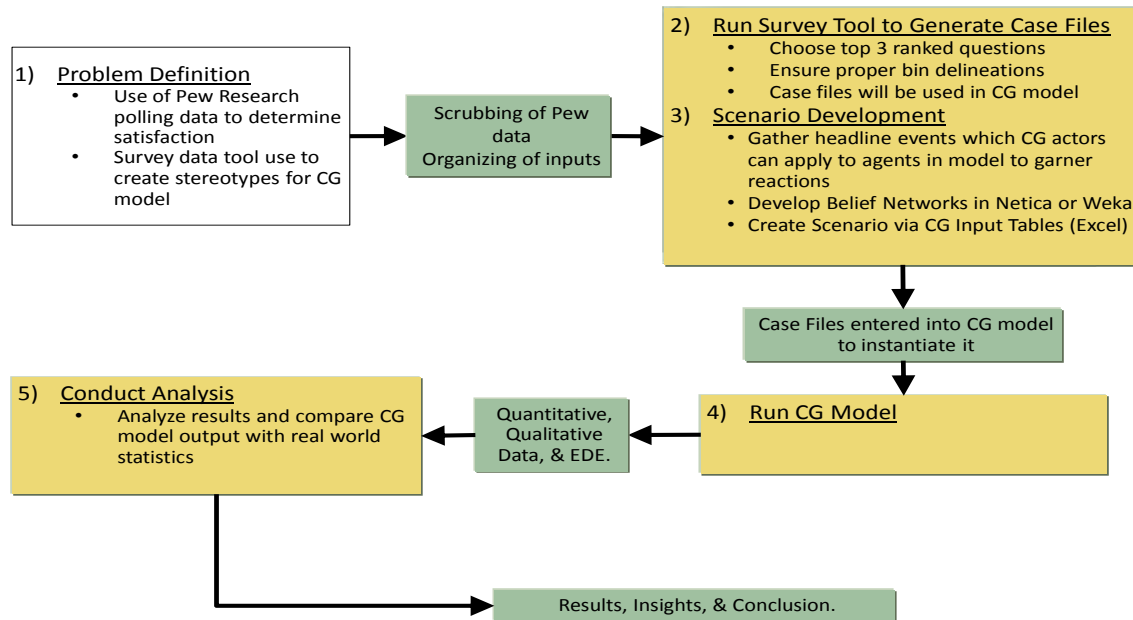


Figure 14. Research methodology for Survey Data Development tool

### 1. Gathering the Survey Data

Figure 14 illustrates the method used to evaluate the survey data development tool. As mentioned above, this tool takes survey data from one or more sources, generates stereotypes based on the survey results, and produces case files that are fed into a social simulation, such as the Cultural Geography model. For this part of this research, polling data was gathered from the Pew Research Center for the People and the Press (<http://people-press.org/>). The surveys spanned nine months from January to September 2010, and eight total surveys were used (one each from January and February, two from March, and one each from April, May, August, and September). See Appendix A for a sample Pew survey. The targeted question examined in all the surveys asked participants about their satisfaction with the state of the nation (at the time the survey was taken), and respondents were allowed to answer “Satisfied,” “Dissatisfied,” or “Don’t

Know/Refuse.” Figure 15 shows the responses of participants for the satisfaction question over 20 years of Pew Research polling. Figure 16 shows the satisfaction levels for the first nine months of 2010.

The original survey data did not include dynamic inputs. Dynamic inputs are significant as they act as catalysts to affect change in population responses. For this research, three dynamic factors were added to the survey data. These were the national unemployment rate, the economic confidence index, and oil prices. The former two are explained below in Section III.B.5. The original data for these factors were pulled from their respective websites (see III.B.5) and entered into Excel. Once in Excel, the data were trimmed to match the dates of the survey data, and then they were put into the master survey in JMP Pro 9. New columns were created in master survey and the factors were copied over from Excel.

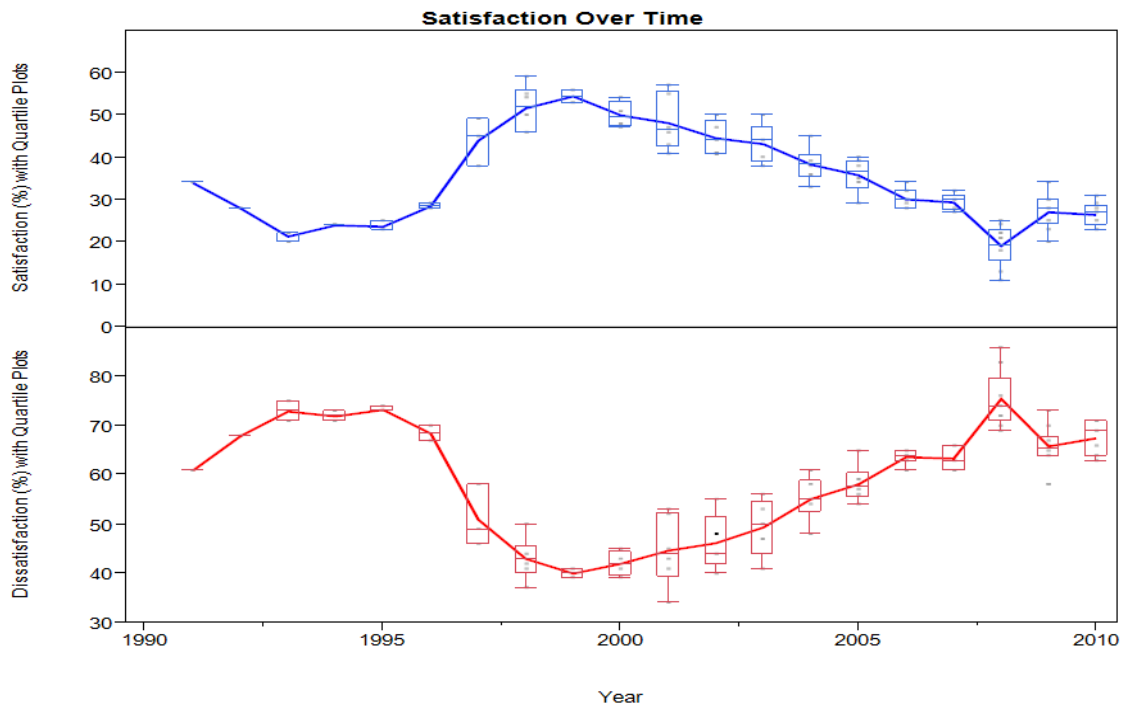


Figure 15. Twenty year running track of satisfaction with current state of the nation (From Pew, 2010)

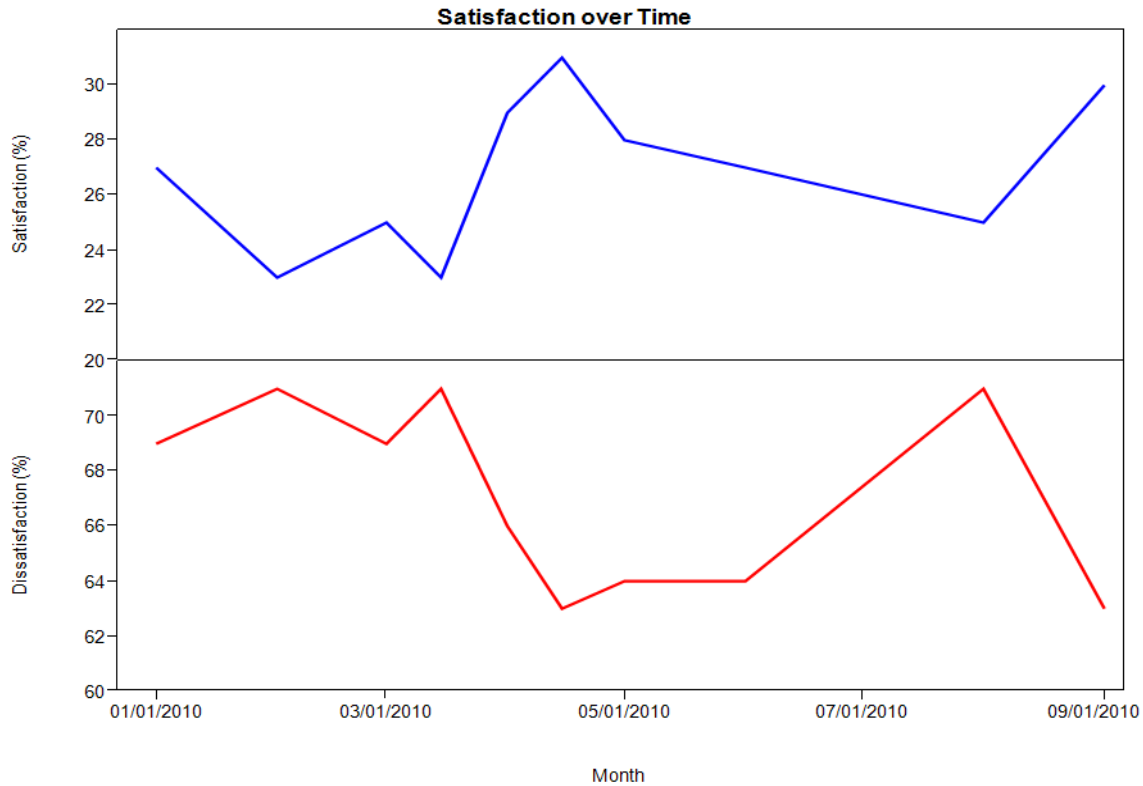


Figure 16. Snapshot of results for Pew satisfaction question for 2010

## 2. Scrubbing and Cleaning the Data

Prior to being able to use the surveys in the tool, the surveys had to be concatenated into one combined survey. This was done using JMP Pro 9 statistical software. Once all the surveys were combined into one large survey, similar question then had to be combined and consistently formatted. For example, the question of age was asked of the respondents in all surveys, but it was not always asked in the same manner for each survey. One survey would ask for a specific age, while another had the respondents give their age based on a range of ages. The principle of the least common denominator was applied, and question results were put into a form that covered all possible responses. In the age example, ages were finally put into ranges that could be interpreted by the tool. Finally, to remove irrelevant information from the surveys, respondents who refused to answer or did not know the answer to a major question and term (answer to satisfaction question, statements regarding race, income, etc.) were removed from the data sets. This was accomplished in JMP Pro 9 by creating subsets of

surveys, and eliminating those respondents as they were identified. Only respondents who answered “satisfied” or “dissatisfied” to the question of national satisfaction were used in the survey data.

### **3. Establishing the Demographics and Organizing the Data**

In order to develop stereotypes for use in the CG model, a series of consistent, persistent questions that were present in all the surveys and had a possible chance of directly relating to the question of satisfaction had to be used. After that, questions that scored relevant to satisfaction and the dynamic national events and national sentiment could be incorporated. The obvious starting location was the demographics of the respondents, as these pieces of information were present in all the surveys and seemed like a logical starting point for predicting satisfaction. The demographics used for this part of the research were age, community type, education level, income, political party affiliation, and race. Each demographic is explained below. It is significant to note that several of the demographics were simplified (compressed or significantly reduced in number) in order to reduce the overall number of stereotypes entered into the CG model. This was done to reduce run time and memory requirements as well as reduce the overall number of case files generated. Table 5 summarizes the demographics chosen.

#### ***a. Age***

As mentioned previously, age was asked of the respondents in a variety of manners. With the variation in response, it was thought best to combine the ages into three broad reaching ranges that would adequately cover the spectrum of respondents and provide useful information into the survey tool. The three ranges were “young” (less than 36 years), “middle age” (36–55), and “senior” (older than 55).

#### ***b. Community Type***

Another characteristic deemed significant to the satisfaction of a respondent was the location of that individual within the country. Pew asked several questions about location and provided many aspects of respondent residence that could have been used in this research. Respondents were asked their zip code, and Pew was

able to determine their state, region of the country, and the type of community the respondents lived in. Community type was chosen based primarily on the fact that it only contained three choices and would reduce the number of stereotypes developed. Community type was determined by Pew based on zip code, and the three types were Urban, Rural, and Suburban.

*c. Education Level*

A respondent's education level could also shed some perspective on his/her satisfaction with the state of the nation. Thus, four categories of education level were constructed from the Pew results. These categories were ~~LT~~HighSchool"—less than a high school diploma; ~~H~~HighSchool"—a high school diploma; ~~S~~SomeCollege"—some college; and ~~GT~~College"—a college degree or higher.

*d. Income*

A respondent's income appears that it would directly affect his/her satisfaction with the nation. For this study, four income levels were created from the nine income levels from which Pew divided the respondents. Pew's original income brackets were ~~L~~ess than \$10,000"; ~~10~~ to under \$20,000"; ~~20~~ to under \$30,000"; ~~30~~ to under \$40,0000"; ~~40~~ to under \$50,000"; ~~50~~ to under \$75,000"; ~~75~~ to under \$100,000"; ~~100~~ to under \$150,000"; and ~~\$~~150,000 or more." These were retracted into the following four brackets: ~~<~~40K"; ~~40~~–75K"; ~~75~~–150K"; and ~~>~~150K."

*e. Political Party Affiliation*

It appeared obvious that a classification such as political party affiliation would certainly correlate to how someone feels about the state of the nation. While there were several parties available to Pew respondents, this thesis used respondents who classified themselves as Democrats, Republicans, and Independents.

*f. Race*

The final demographic extracted from the survey is race. There were numerous races as options for respondent selection in the survey (black, white, Hispanic,

Asian, Native American, and other). In the same vein as other demographics, race was simply divided into three categories (black, white, and other).

Table 5. Final demographic bins of respondents for survey tool

CommunityType	PoliticalAffiliation	Age	Race	Education	Income
Urban	Democrat	Young	White	GTCollege	<40K
Rural	Republican	MiddleAge	Black	SomeCollege	40-75K
Suburban	Independent	Senior	Other	HighSchool	75-150K
				LTHighSchool	>150K

#### 4. Adding Additional Responses to Account for Multiple Questions

As mentioned above, a question might be asked in several surveys in different ways. This was true of the question on how respondents viewed President Obama and their satisfaction with the job he has done as President. This particular question was initially deemed significant to overall national satisfaction, and thus it needed to be included in all the surveys. (It was later discovered that the tool did not score Obama's rating as significant.) Across the year of surveys, that question was asked in three different ways. The three questions and possible responses were:

- *Do you approve or disapprove of the way Barack Obama is handling his job as President?*
  - Approve, Disapprove
- *Would you say your overall opinion of Barack Obama is very favorable, mostly favorable, mostly unfavorable, or very unfavorable?*
  - Very favorable, mostly favorable, mostly unfavorable, very unfavorable
- *Now I'd like to ask your impression of some groups and individuals. Would you say Barack Obama is doing an excellent, good, only fair, or a poor job?*
  - Excellent, good, only fair, poor

In order to get a response to these questions across all surveys, a numerical index was established that ranked and took into account the responses from the participants. The index, titled ObamaIndex in the tested surveys, consisted of two scores: 1 or 2. A

score of 1 was designated if a response was “Disapprove,” “Mostly unfavorable,” “Very unfavorable,” “Only fair,” or “Poor.” A 2 was given otherwise. This new index, accounting for all the responses across all eight months of surveys, was included in the tool testing data sets.

## **5. Establishing the Dynamic Inputs to the Surveys**

Once the demographic information was created and formatted and similar-topic questions were concatenated and indexed, the dynamic input to the surveys was gathered, processed, and placed into the master survey. As mentioned above, three dynamic inputs, called indicators, were added to the surveys. These were the national unemployment level (%), the economic confidence index, and oil prices (\$ per barrel). Figure 17 shows the plots of consumer economic confidence and national unemployment across 2010. Figure 18 is the confidence index reversed (negative numbers positive, thus higher number is worse index). This figure is shown so that the trends for a worse economic confidence and increasing unemployment would go in the same direction, in an effort to illustrate any potential relationship between the two indexes. The data was taken from the Gallup website (<http://www.gallup.com/poll/economy.aspx>) and the graphs were produced using JMP Pro 9 software. Quantshare.com defines the Economic Confidence Index as

...a combination of the Economic Outlook and Economic Conditions measures created by Gallup. The U.S. Economic conditions results from a rate given by Americans regarding today's economic conditions in the country, while the U.S. Economic outlook results from Americans opinion on whether economic conditions in the U.S. are getting better or worse. Both measures are obtained from telephone interviews with more than 3000 U.S. citizens. The Economic Confidence Index is simply the sum of these two measures (Economic conditions and economic outlook).

A negative value in the Economic Confidence Index indicates that more Americans have negative than positive sentiments regarding the U.S. economy. This is the case since January 2008 and the index still remains in a negative territory. As of November 15, 2010, the Economic Confidence Index value is -28; a decrease of 4 points since the beginning of the month (-24) and 8 points since the beginning of the year (-20). (Bug Man, 2011)

The oil price data was gathered from the West Texas Research Group (WTRG) website (<http://www.wtrg.com/daily/oilandgasspot.html>) and spanned over 14 years of price tracking from 1997 through March, 2011. The periods of interest for this thesis were all of 2010 (survey tool application) and the first six months of 2011 (STANLEY application). Figures 19 and 20 show the oil prices over all 14 years. Figure 20 is the West Texas Intermediate (WTI) Cushing oil price and the price index used in this thesis. Figure 21 shows the prices for 2010.

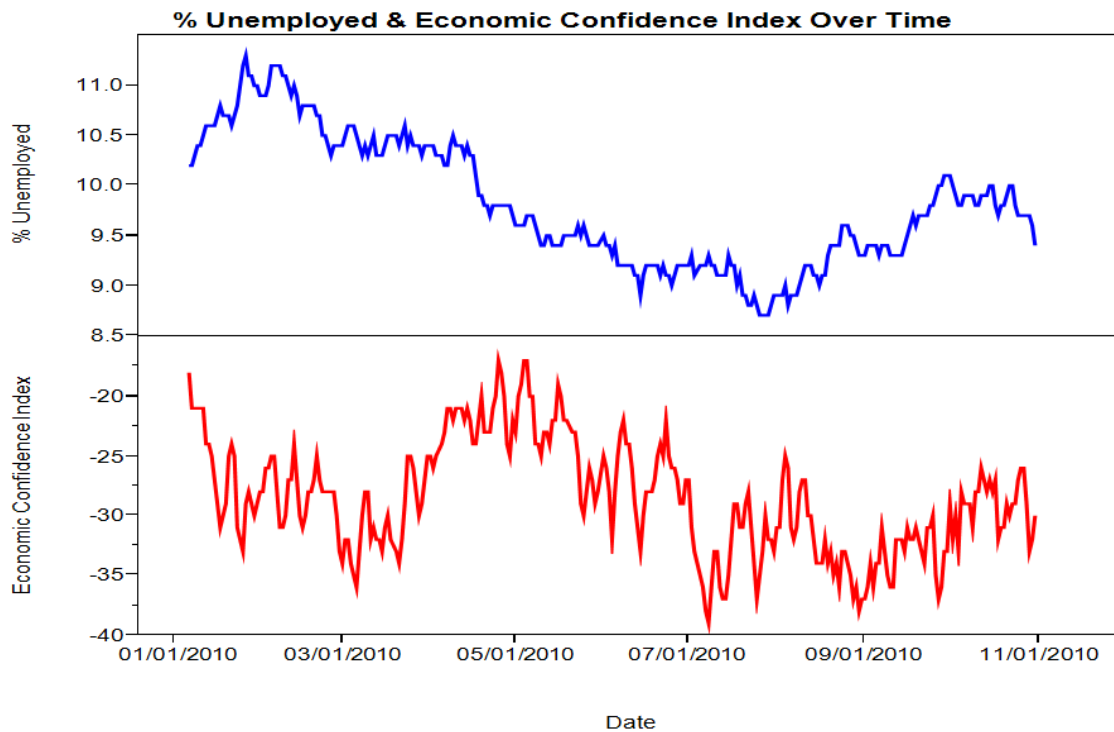


Figure 17. Unemployment rate and economic confidence index for 2010

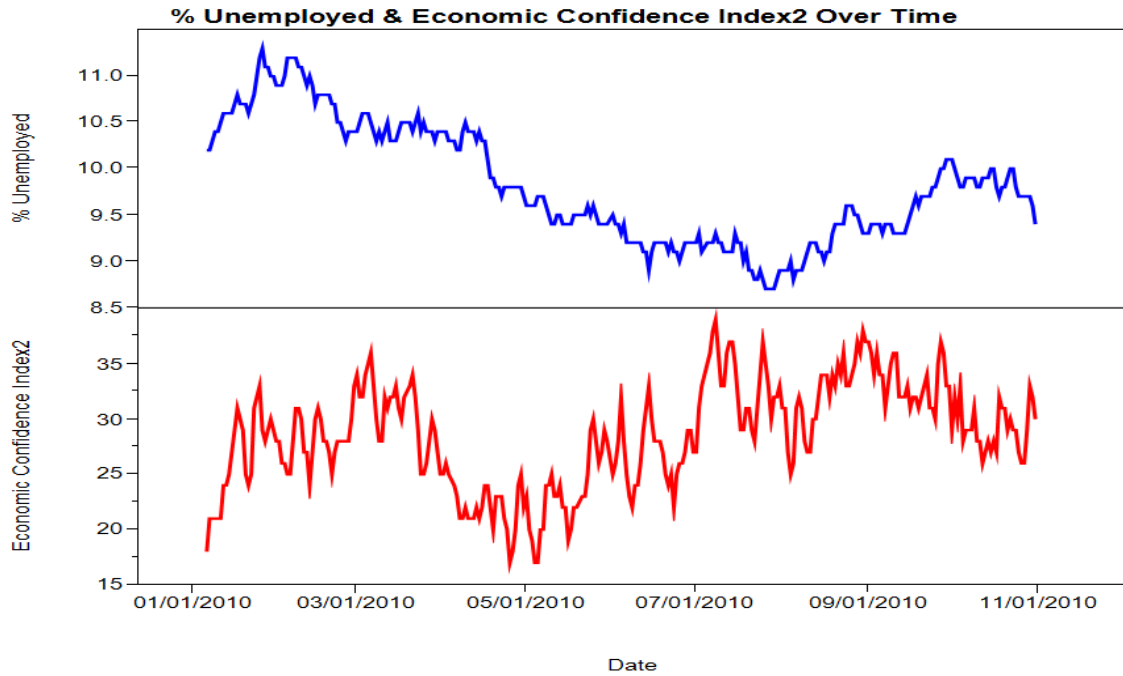


Figure 18. Unemployment rate and economic confidence index for 2010, reversed. Higher confidence index indicates worse consumer outlook.

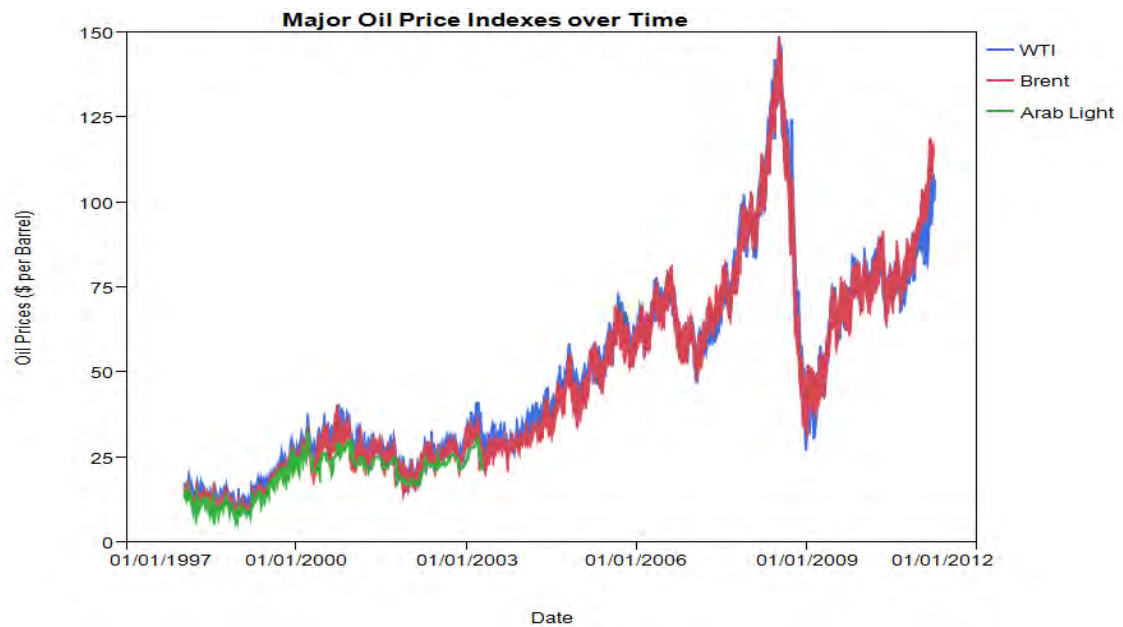


Figure 19. Oil Price data from 1997 through March 2011. Three major oil price indexes shown (best viewed in color)

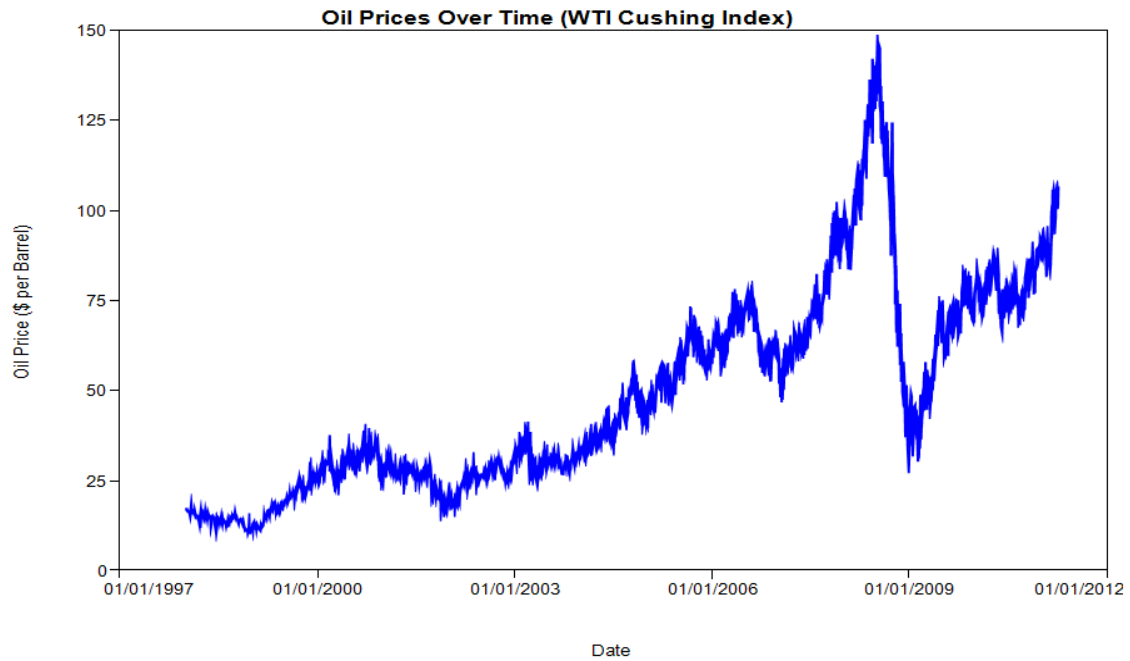


Figure 20. WTI Cushing (index used in thesis) Oil prices from 1997–March 2011

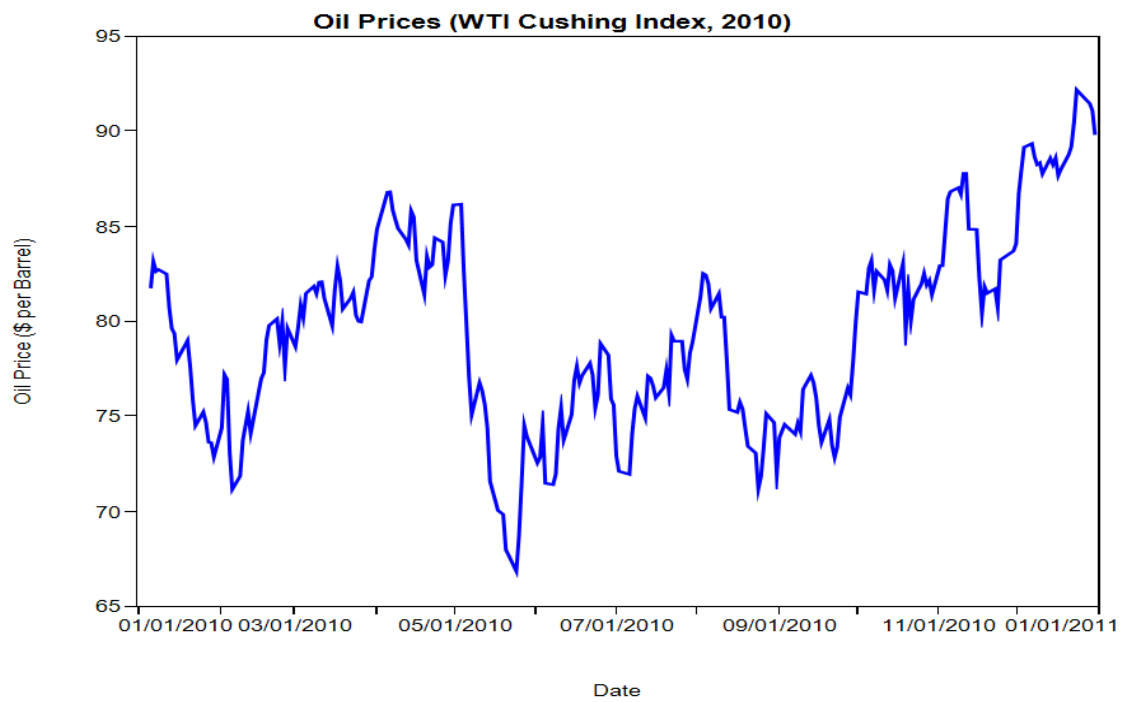


Figure 21. WTI Cushing Price Index for 2010 (\$)

As a tracking mechanism for the model, categorical referents were created from the continuous unemployment, economic confidence, and oil price data. The referents were created simply by assigning a value of “NegTrend” to the initial value in the survey, and if the index then trended in a manner that would have a negative impact on society (i.e., higher oil prices or unemployment). If the index trended toward a positive impact, the value would change to “PosTrend.”

## **6. Culling the Surveys: Removing Insignificant Questions**

Once the concatenated survey was filled with dynamic data, and all the important demographic data was included for every respondent, the survey was culled of questions of lower significant effect on the question of national satisfaction. This was accomplished using the partition feature in JMP Pro 9. From the JMP website, with Y representing the question of national satisfaction and X representing the responses

JMP’s partition platform enables users to systematically analyze large data sets to discover unsuspected or unknown relationships. JMP uses visualization to create a successive tree of partitions according to a relationship between the X and Y variables. It finds a set of cuts or groupings of X values that best predict a Y value by exhaustively searching all possible cuts or groupings, recursively forming a tree of decision rules until the desired fit is reached. Through the use of visualization and recursive portioning, JMP makes data mining techniques accessible (“Effective Data Mining,” 2011).

Figure 22 is an example of the JMP Pro 9 output used to make decisions on which questions to remove from the survey. Early in the process, while the survey contained over 150 questions (including demographics and dynamic input), questions with zero splits were removed. The partition model was run with K-fold validation set at 100 folds. Higher folds were attempted, up to the maximum K value, but the results were not significant enough to justify the time spent running the model. Once the number of remaining questions was reduced to 35, JMP’s  $G^2$  rating was used to remove low significance questions.

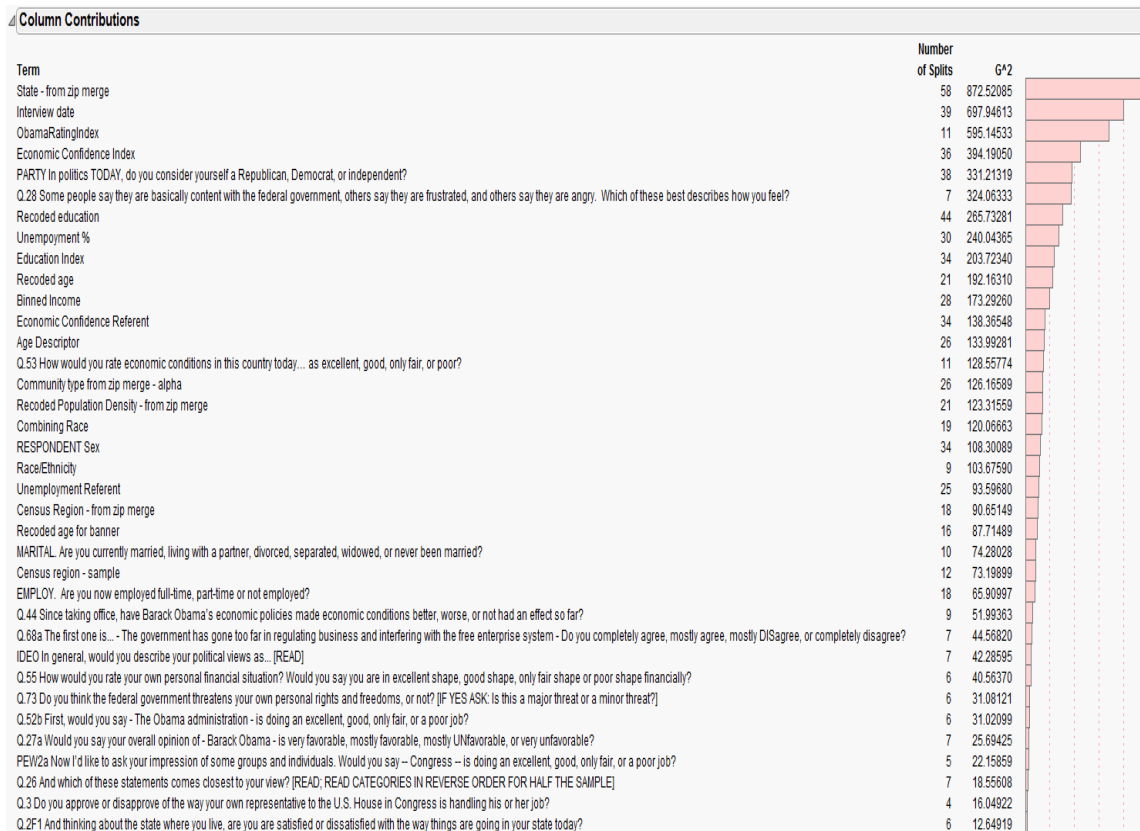


Figure 22. Sample JMP Pro 9 output used for culling survey questions

Ultimately, 23 questions (not including all demographic data and dynamic input; 48 total questions including all that data) were deemed as significant to the determination of national satisfaction among the respondents. It was left to the survey data tool to further reduce the number of questions by ranking them in significance to the satisfaction question.

### 7a. Running the Unaltered Tool: Generating Case Files Using Modified Enumerated Responses

As discussed previously, the initial issues with the tools were discovered through trial and error. In the initial tool runs, it was immediately noticed that the tool would not take string responses. While awaiting authorization for Mr. Sullivan to alter the tool's original code, it was determined to continue with the application of the tool using altered surveys. Essentially, all survey questions that had string answers needed to have those

responses turned into positive integers that corresponded to the response. Table 6 shows the demographic data and the enumerated values for that data used in the initial tool runs.

Table 6. Demographics with enumerated values to the right

CommunityType	PoliticalAffiliation	Age	Race	Education	Income
Urban -- 1	Democrat -- 1	Young -- 1	White -- 1	LTHighSchool -- 1	<40K -- 1
Rural -- 2	Republican -- 2	MiddleAge -- 2	Black -- 2	HighSchool -- 2	40-75K -- 2
Suburban -- 3	Independent --3	Senior -- 3	Other -- 3	SomeCollege -- 3	75-150K -- 3
				GTCollege -- 4	>150K -- 4

With the demographic levels established and the surveys transformed into the enumerated forms of the demographics, they were saved as .csv files so they would be compatible with the tool. Training and test set surveys were then chosen. The training set included the surveys from January, February and the first set of March surveys (these were taken from 10–14 March). The test set contained the remaining surveys.

The ultimate purpose of the tool is to generate case files for response prediction in the CG model. Case files were created from both the training and test sets. The first set of case files created contained only the demographics explained above in A.5.a-f. The second added the categorical dynamic conditions (an unemployment referent, economic confidence referent, and an oil price referent). The third set was created from the second by including a satisfaction score for President Obama that was gleaned from the surveys. The fourth set of files included the top three responses (in addition to the demographics, dynamic categories, and the Obama Index) as determined by the Chi-squared calculations performed by the tool. The top three questions returned were:

- *Do you [approve/disapprove] very strongly, or not so strongly (of the way Barack Obama is handling his job as President)?*
  - This question was labeled in the tested survey as Q2aPresSat with possible responses of Very strongly and Not so strongly (note: this question was removed later in the process. See section III.A.8 below for explanation.)

- *How much effect do you think the federal government's activities—such as the laws passed and so on—have on your day-to-day life? Do they have a great effect, some effect, or no effect?*
  - This was labeled Q20GovEffect in the tested survey data with responses of None (no effect), Some (some effect), and Good (great effect)
- *Considering what you get from the federal government, do you think you pay more than your fair share of taxes, less than your fair share, or about the right amount?*
  - This was labeled Q23TaxBurden with responses of Less (respondent gets less than what he pays for), AboutRight (respondent receives right amount), and More (respondent receives more than he pays for).

Of note, the time required to run the tool varied greatly depending on the number of case files generated. For the demographic only run, 1296 files were created, while over 380,000 were created for the top three question run. The time to run the smaller test was about 10 minutes. The time to execute the top three run exceeded four to six hours, depending on network strength and the size and number of files generated. Of significant use in the CG model was the fourth run that included the top three rated questions. The original plan for the tool test included using this set of case files for use in the CG model. Then the code was rewritten to allow strings, and the case files were regenerated using the original survey answers. Additional results from this initial testing resulted in further modifications to the tool as well in order to increase efficiency, reduce tool run times, and allow the user to specifically select stereotype terms and other data.

#### **7b. Running the Modified Tool: Generating Case Files Using Original Responses**

The code was not only rewritten to include string responses, but it was also modified so that the user could select those terms that were to be included as stereotypes in the generation of case files and other data that would be included in the case files. This was a significant modification, as it allowed the user to provide as much information regarding other factors of interest (in this case the dynamic inputs and top three

questions) to the CG model through the case files while minimizing the size and number of case files generated (this also resulted in a significant decrease in the time required to generate the case files). Figure 23 illustrates the new user interface for selecting these terms. As an example, in the figure, “Education” and “Income” would be part of the stereotypes developed in the generation of case files, while data from OilPriceIndex, EconConfRef, Q13, and Q20 would be included in those case files as well.

**Demographic Stereotype Definition**

**OilPriceIndex Category**  
☒ Use each value to designate a demographic bin ☐ Stereotype  
 Number of Demographic Bins: 6  

Long Name	Short Name	Range (Low)	Range (High)
OilPriceIndex...	OINegTrend	NegTrend	NegTrend
OilPriceIndex...	OIPoSSTrend	PosTrend	PosTrend

**Education Category**  
☐ Use each value to designate a demographic bin ☒ Stereotype  
 Number of Demographic Bins: 4  

Long Name	Short Name	Range (Low)	Range (High)
EducationLT...	E4	GI College	GI College
EducationLT...	E1	LTHighSchool	LTHighSchool

**EconConfRef Category**  
☒ Use each value to designate a demographic bin ☐ Stereotype  
 Number of Demographic Bins: 6  

Long Name	Short Name	Range (Low)	Range (High)
EconConfRef...	ECNegTrend	NegTrend	NegTrend
EconConfRef...	ECPosTrend	PosTrend	PosTrend

**Income Category**  
☐ Use each value to designate a demographic bin ☒ Stereotype  
 Number of Demographic Bins: 4  

Long Name	Short Name	Range (Low)	Range (High)
IncomeI1	I1	I1	I1
IncomeI4	I4	I4	I4

**Q13PerFinances Category**  
☒ Use each value to designate a demographic bin ☐ Stereotype  
 Number of Demographic Bins: 6  

Long Name	Short Name	Range (Low)	Range (High)
Q13PerFinan...	Q1		
Q13PerFinan...	Q1Fair	Fair	Fair

**Q20GovEffect Category**  
☒ Use each value to designate a demographic bin ☐ Stereotype  
 Number of Demographic Bins: 6

Figure 23. Improved survey tool GUI. Includes ability to select stereotypes

Once the tool’s code was modified by Mr. Greg Sullivan, string responses could be included in the surveys for analysis. With this development, it was necessary to go back and regenerate case files. Similar processes as above were used to generate these files. The demographic data remained the same (no longer enumerated, however). Some modifications were made to the surveys’ allowable answers, as some were phrases or

sentences. To simplify the processing done by the tool and hopefully shorten the case file generation time, these sentences and phrases were condensed into one or two word answers. For example, one of the top three questions was *“Considering what you get from the federal government, do you think you pay more than your fair share of taxes, less than your fair share, or about the right amount?”* The original responses, as recorded by Pew were 1) fair share, 2) less than fair share, and 3) about the right amount. These responses were shortened to 1) Fair, 2) Less, and 3) aboutRight.

The time required to run the case files from the updated tool took only about three minutes to generate 1,296 case files (the number of combinations of the demographic divisions).

## **8. Creating the Bayesian Network shell**

Once the case files were completed, the Bayesian network shell needed to be developed for later input into the CG model. CG uses a software package like Netica or Weka for the importation of Bayesian Networks. Netica is an open source program developed by Norsys Software Corporation. The purpose of the software is to develop belief networks and influence diagrams that can be transferred to other networks or models for problem solving and analysis (*“Netica application,”* 2011). Weka is a similar open source program developed by the University of Waikato built to mimic the abilities of IBM’s SPSS statistical analysis software.

### ***a. Adjusting the Survey Data to Meet Netica Restrictions***

When the case files generated in III.B.7b above were put into Netica in order to generate belief network nodes, a few issues were discovered. First, Netica separates responses with spaces. Any node (nodes are based on column names, or questions, in the case files) whose name has a space was broken down into two nodes. For example, the demographic *“Political Party”* was split into two nodes, *“Political”* and *“Party.”* Second, Netica could not handle nodes whose various states (states were the responses for each question, or node, in the survey data) began with numbers. It needs non-numeric characters as the start of state names.

These issues needed to be handled prior to continuing with the model implementation. The spaces in the surveys were easily removed using Excel's find and replace function. Fortunately, the only applicable question that included numeric responses was the question of income. To correct this, the income levels were given a short name that corresponded to the stereotype described above. These names were:

- I1 → <40K
- I2 → 40–75K
- I3 → 75–150K
- I4 → >150K

***b. Rerunning the Case Files with Corrections to Surveys***

Finally, in order to proceed, the case files needed to be regenerated using the new and improved survey data. Beneficially, with the opportunity to rerun the tool, the surveys were reexamined, and several prudent changes were made. In addition to the changes made to generate new case files, the question Q2a that was rated as a top three question above was reevaluated and converted into a new question. Originally, the question asked to what degree a respondent was satisfied or dissatisfied with the job of President Obama. It was felt that the responses from that question, which were “Very strongly” and “Not so strongly,” by themselves, were not applicable to the overall question of national satisfaction. They needed to be included in their context, so a new question was created to combine whether a respondent was satisfied or dissatisfied with President Obama and to what extent. For example, if a respondent was satisfied with the President but not very strongly, then the new question, called Q2Combo, would be answered “SatisfiedNotSoStrongly.” Also, it was determined to run the training data and the entire survey data in the tool and generate case files for each. This was done for purely comparative purposes regarding the ultimate output from the CG model

After removing Q2a and replacing it with Q2Combo, the survey tool was run with both the modified training data set and the entire modified survey data set. New top three questions emerged in both. For the training set, the top three questions were:

- Q20GovEffect: *How much effect do you think the federal government’s activities—such as the laws passed and so on—have on your day-to-day life? Do they have a great effect, some effect, or no effect?*
- Q23TaxBurden: *Considering what you get from the federal government, do you think you pay more than your fair share of taxes, less than your fair share, or about the right amount?*
- *What do you personally feel is the bigger problem with government? Government has the wrong priorities, OR government has the right priorities but runs programs inefficiently?*
  - This was labeled Q22GovProblem with responses of Neither (problem is not with priorities or inefficiency), Priorities, Inefficient, and Both (both priorities and inefficiency are problems).

For the entire survey, the top three questions were:

- Q20GovEffect: *How much effect do you think the federal government’s activities—such as the laws passed and so on—have on your day-to-day life? Do they have a great effect, some effect, or no effect?*
- *How would you rate your own personal financial situation? Would you say you are in excellent shape, good shape, only fair shape or poor shape financially?*
  - This was labeled Q13PerFinances with responses indicating respondents’ personal financial situation as Poor, Fair, Good, or Excellent.
- Q23TaxBurden: *Considering what you get from the federal government, do you think you pay more than your fair share of taxes, less than your fair share, or about the right amount?*

As above, the case files were generated using these questions, as well as demographics and dynamic inputs to the survey data. Figure 24 is an illustration of what a Bayesian network generated in Netica would resemble using the entire survey data; Figure 25 shows the network using the results from the training data set.

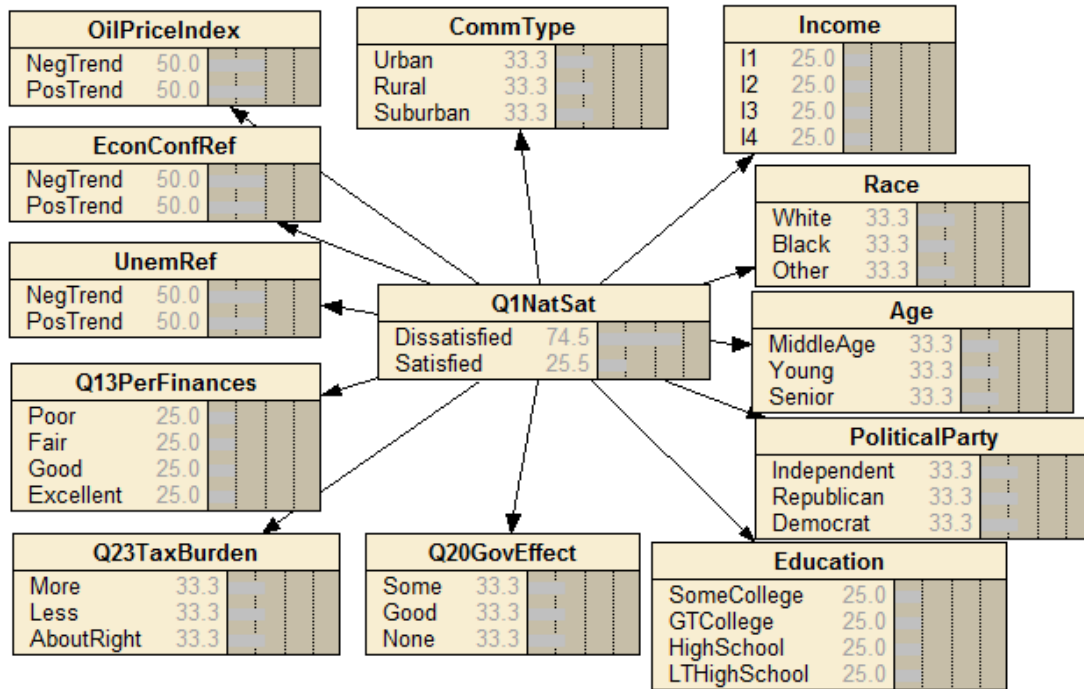


Figure 24. Netica generated Bayesian network illustration given demographics, coded indexes, and top three questions for entire survey

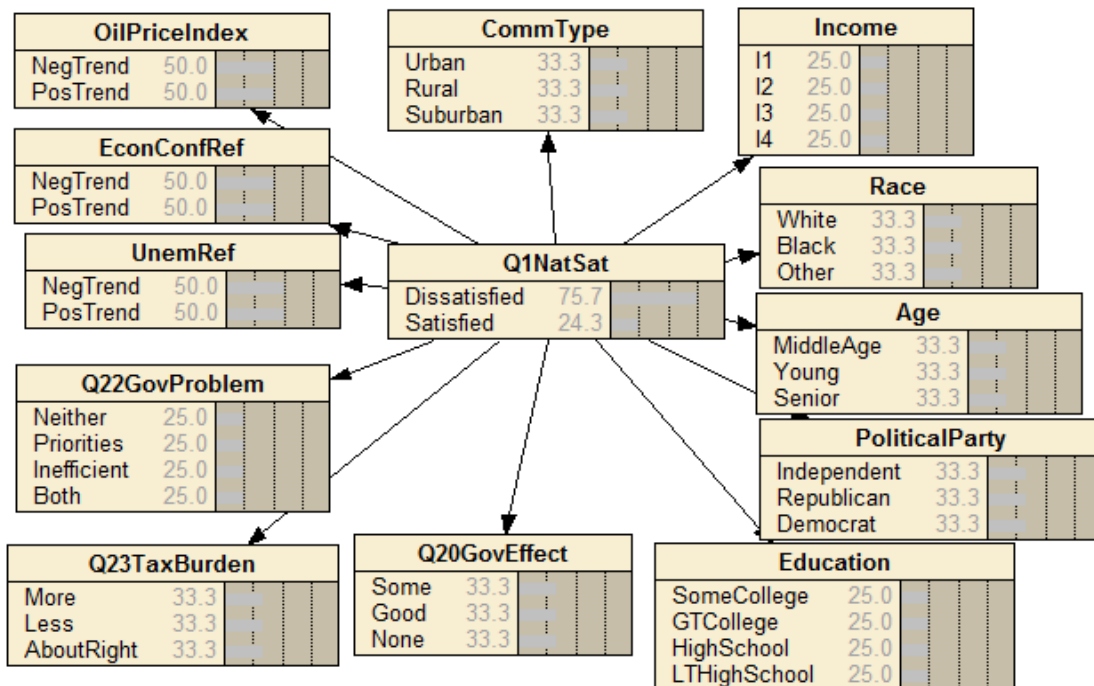


Figure 25. Netica generated Bayesian network for Training Survey

## **9. Scenario Development—Preparing the Case Files and the CG Input Tables for Use in the Model**

In order to satisfactorily test the tool's effectiveness and its ability to generate sufficient case files, appropriate scenarios needed to be developed that could truly determine the worthiness of the tool's ability to instantiate the CG model. The actual creation and instantiation of the input tables for CG is beyond the scope of this thesis, but there are a few items that need highlighting.

### ***a. Survey Tool Scenario***

The initial input to the CG model using the case files was based on a 20 agent setup. These agents were “hand-made,” meaning that they were specifically given certain stances on the top three questions from the survey tool results using the entire survey dataset and were deposited equally into one of four regions within the CG scenario. These regions were based on actual survey regions (regions were denoted in the survey for each respondent). These regions were South, Northeast, Midwest, and West. The agents were developed in a manner that instantiating and running the model would lead to simpler debugging. Essentially, this first iteration of the model run was conducted purely to ensure the proper execution of the model including proper injection of beliefs via the established Bayesian networks, sufficient communication between the agents, and noticeable changing of issue stances over time.

Subsequent tests of the CG model included all the case files generated from the survey data. There were 1,296 case files generated based on the demographics of the respondents. Initially, for each of these case files, agents were developed and placed in the regions in comparable proportions to the actual survey demographics. For the South region, two of each type agent was created; and one of each from the Midwest, West, and Northeast. This resulted in a total of 6,482 agents in the CG model (at the time of this research, this was the most agents ever built in the model!). However, this amount of agents proved to be too time consuming, as the model bogged down as all the possible connections between the agents (homophily) were created. After several attempts to run the CG model with this many agents, it was decided, due to numerous system crashes and

excessive time requirements (nearly 20 hours to run only seven days of a 249-day scenario), to reduce the number of agents to match the number of stereotypes (1,296). All the agents were placed in the same region. This was done for the sake of improved efficiency to allow more model runs.

## **10. Methods of Analysis**

Several CG scenarios were developed and run. Described in more detail in IV.A, these scenarios were incremental and involved a smaller agent run, a full run with the training survey data (10 row minimums for the case files), and full runs with the complete data sets (one without and one with the 10 row minimums). The results of the CG model were analyzed to see if the Survey Data Development Tool generated case files of sufficient quality such that the CG model's output similarly compared to real world data. In this case, CG's results were compared to the actual survey data results gathered by the Pew Research Center. To make this comparison, the satisfaction results from CG, which initially represented an agent's level of satisfaction, on a scale of zero to one, had to be transformed into the number of agents who responded "~~satisfied~~" to the question of national satisfaction, thus matching the format of the Pew results used herein. This was accomplished by designating agents with a satisfaction level greater than 50% (0.50) as "~~satisfied~~," and those below as "~~dissatisfied~~." This allowed for simple math and spreadsheet manipulation functions within JMP Pro 9.0 Pro to compute the percentage of agents who were "~~satisfied~~." The main indicator of validated output was the direction of trends in satisfaction over time and how these directions compared to actual Pew data. Additionally, tests were done comparing CG output generated by case file inputs that had no minimum row restriction with output where the case files had a minimum row restriction of 10 rows. This was done to determine if there was a significant difference in output generated by the different case file types. Section IV.A describes the method of analysis and results from the testing of the survey tool.

## C. STANLEY RESEARCH METHODS

This section provides a brief overview of the methodology that will be used in analyzing and evaluating the text analysis tool.

### Methodology-STANLEY

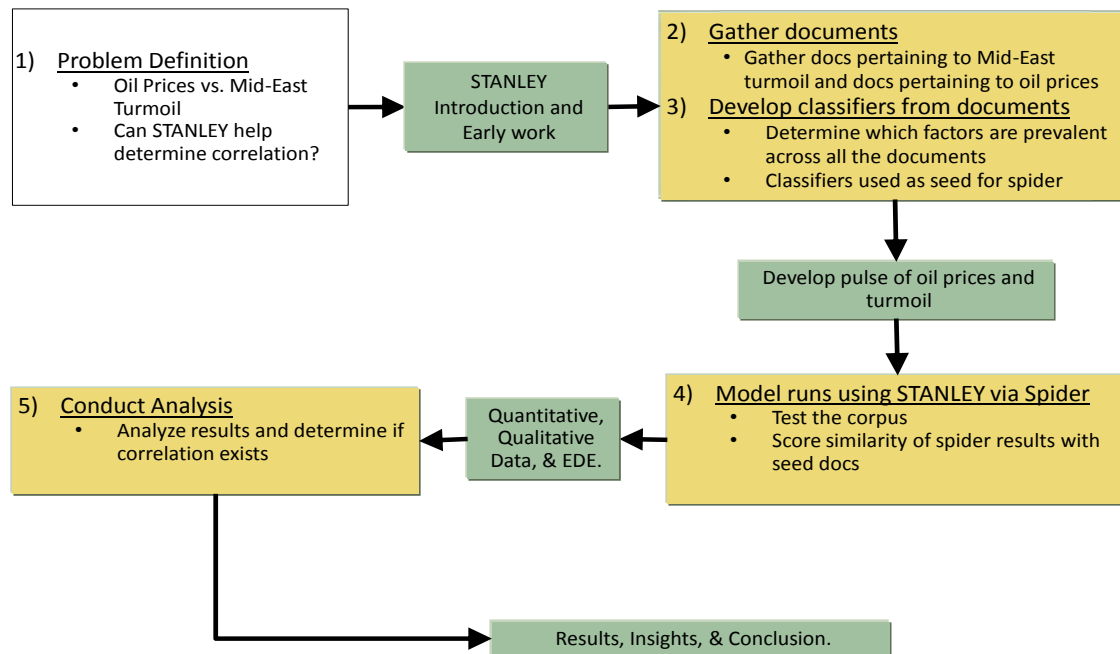


Figure 26. Research Methodology for Sandia's text analysis tool

The methodology for the STANLEY tool analysis was similar in overall set up as the survey tool, but there were some significant differences. It extends the research done by McKaughan et al (2011) by automating the document corpus generation via a web spider and looking for correlation between the documents and a real world situation by comparing document similarities. The overarching goal of this portion of the thesis was to see if STANLEY could determine if there was a correlation between the sentiment expressed in the media regarding Middle East turmoil and changing U.S. oil prices. Particularly, in the early months of 2011, the Middle East witnessed significant civil uprising and protests with attempts at political reform, with varying degrees of success. During this same period of time, prices for oil increased over 33% (Ross, 2011). Figure

27 tracks oil prices throughout the early months of 2011. Figure 28 plots the oil prices specifically for early 2011. This thesis attempted to see if STANLEY could identify a correlation between media coverage of Middle East unrest and oil prices.

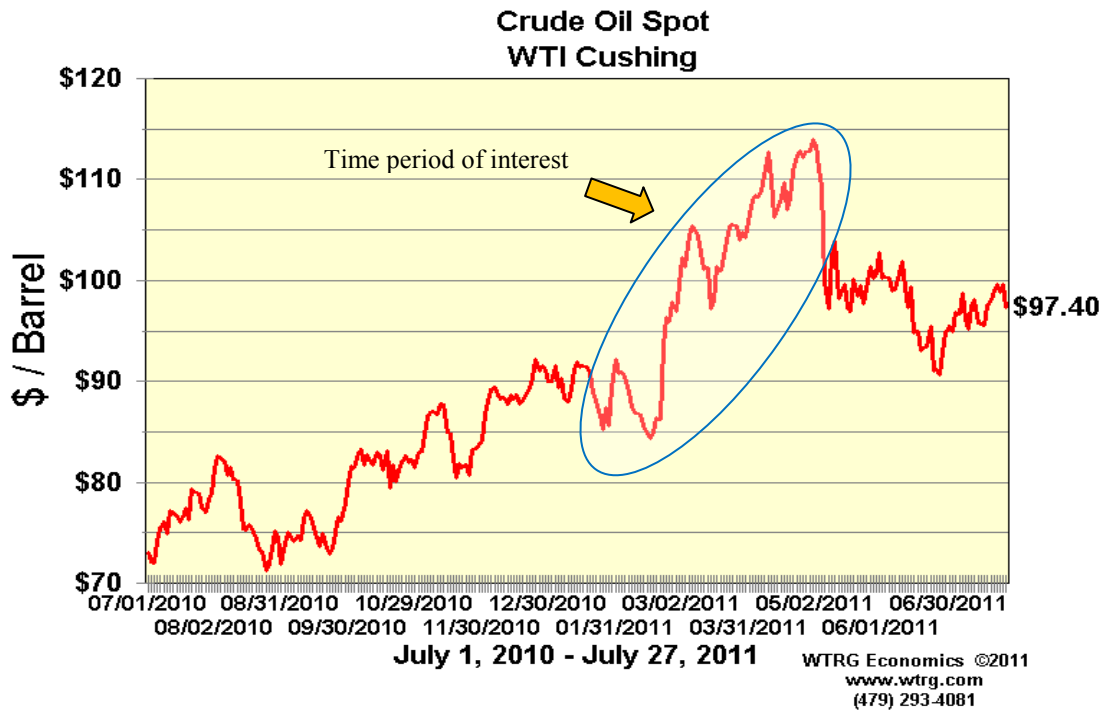


Figure 27. WTI Cushing Price Index  
(From <http://www.wtrg.com/daily/oilandgasspot.html>)

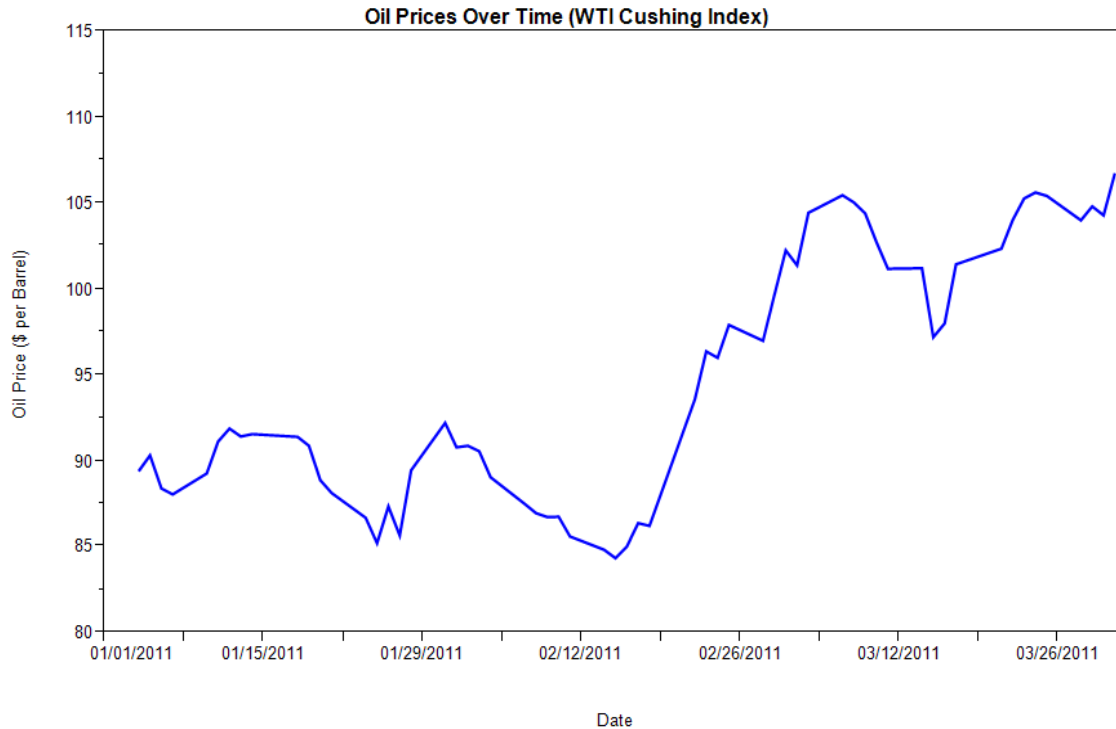


Figure 28. Early 2011 oil prices (WTI index)

This issue defined the problem to solve and led to the following methodology in an attempt to provide an answer.

### 1. Gathering Appropriate and Applicable Documents

The first order of business, after determining the problem, was to gather the documents that STANLEY would analyze. Figures 29 and 30 briefly generalize the process of gathering documents. First, sample documents are gathered that are representative of the content desired among the remaining documents to be collected. Second, the sample documents are used to fetch other similar documents from various sources. While there are several ways to go about this, including manually surfing the Internet and other resources and pulling down seemingly appropriate material, more user-friendly and efficient methods are available. For this thesis, a web spider, or crawler, developed by Sandia National Laboratories was used to scour the Internet and collect documents. Lastly, the retrieved documents are placed in a directory, and matching scores are calculated. These matching scores, using the cosine similarity function

discussed previously, rate the likeness of the retrieved document to the seed documents and were the principle source of analysis for the efficacy of STANLEY.

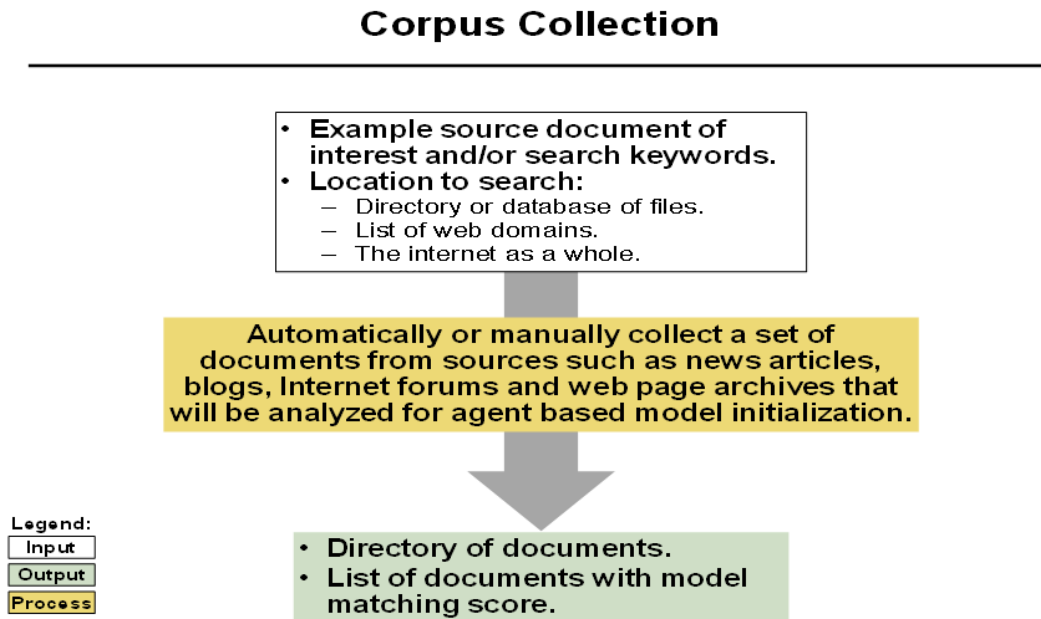


Figure 29. Brief summary of document collection process (After Baez, 2011)

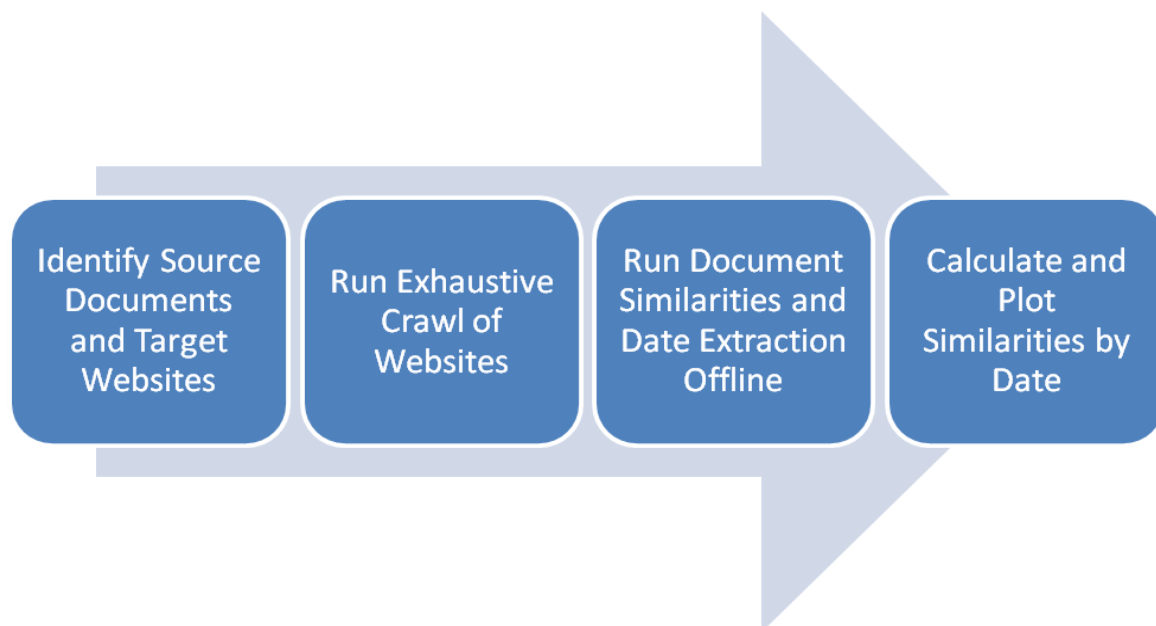


Figure 30. Generalized STANLEY process using spider

*a. Using a Web Spider, or Crawler, to Gather Documents*

A web spider, or crawler, is a computer program that browses the World Wide Web in a methodical, automated, and orderly fashion. In general, it starts with a list of URLs (uniform resource locator) to visit or a body of pre-prepared documents, called the seeds. As the crawler visits these URLs or is pointed to a specific set of URLs with the prepared documents as seeds, it identifies all the hyperlinks in the page and adds them to the list of URLs to visit, called the crawl frontier. URLs from the frontier are recursively visited according to a set of policies (Pant, Srinivasan, & Menczer, 2011).

Web spiders are some of the most powerful and useful tools developed for the Internet, and intentions for their use can be both good and bad. A spider serves one major function—Data Mining. The way a typical spider (like Google) works is by crawling a website one page at a time, gathering and storing the relevant information such as e-mail addresses, meta-tags, hidden form data, URL information, links, etc. The spider then crawls all the links in that page, collecting relevant information in each following page and so on. After a time, the spider has crawled thousands of links and pages gathering bits of information and storing it into a database. This web of paths is where the term “spider” is derived. Figure 4 (section I.B) graphically depicts the process of crawling to gather and process documents. The spider also has the ability to load the downloaded data into a search engine (very similar to Google) which allows the downloaded documents to be searched.

For this research, the spider was seeded with documents manually gleaned from the Internet that were specific in their accounting of the turmoil and uprising in the Middle East in the first third of 2011. The documents used to seed the spider were pieced together from the following websites: <http://www.bbc.co.uk/news/world-12482680>; <http://www.nytimes.com/interactive/world/middleeast/middle-east-hub.html>. Seed documents were also gathered pertaining to the increase in oil prices. The websites used for these documents’ generation were: <http://www.cbsnews.com/stories/2011/04/11/ap/business/main20052645.shtml>; <http://www.heatingoil.com/blog/afternoon-price-check-oil-prices-rally-on-libya-fighting-middle-east-protests0322/>, and <http://neronline.co.uk/economics/what-impact-will-the->

[middle-eastern-protests-have-on-the-world-economy/](#). Finally, a list of URLs was also used to seed the spider. These sites included <http://www.forbes.com>, <http://onlinenews.com>, <http://thestreet.com>, <http://bloomberg.com>, and <http://nytimes.com>. The spider was run at Sandia's laboratory in Albuquerque, NM under the cognizance of Mr. JT McClain, the technical lead for the STANLEY tool.

The spider took on average three to five days to complete a crawl given a particular URL. It produced the following output (generically):

1. A .csv file which lists URLs and their similarity to source documents in rank order
2. A folder with the following directories (this can be quite large)
  - a. DB – Contains an H2 database which contains the following:
    - i. The list of URLs and similarity score
    - ii. Query results from Bing (or Google)
    - iii. Visits to the same URL (the spider does not revisit a URL, but it records when it sees the same URL again)
  - b. XML – Contains XML files which contain text extracted from the documents, as well as other metadata. See Appendix E for example.
  - c. Files – Contains the actual downloaded files

Analyses on the returned results done by the STANLEY tool took an additional one to two days. These calculations included overall similarity scoring, average monthly similarity scores, as well as the median, mode, maximum, and minimum of the similarity scores for a particular month. While the initial spider results were quite large (forbes.com alone initially returned over 70,000 files, greater than 100GB), once a method to remove imagery and extract dates was applied, the numbers were reduced to more manageable levels. For example, the forbes.com results were reduced to 1,800, or roughly 2.5% of the initial total return.

## **2. Developing the Classifiers**

Probably the most significant aspect of effectively utilizing a tool like STANLEY lies in the building of the classification documents, or classifiers. This was noted in the conclusions of the work by McKaughan et al. (2011). They note

The need for classifiers that are orthogonal cannot be stressed enough. The classifiers need to be as specific as possible to the categories they are attempting to represent. As was shown, when the classifiers were well defined...STANLEY was very accurate in classifying the documents. When there existed some ambiguity between texts...the tool was less effective. (pp. 6–7)

Generally, SME input is desired in the development of classification documents. No SMEs were available for the execution of this research, so best guesses were made as to applicable documents with which to seed the spider and which results from the spider would be used in the development of classifiers. The sample documents used to seed the spider are included in Appendix C.

### **3. Setting up and Executing STANLEY**

Unlike earlier tests with STANLEY (McKaughan, Alt, Heath, and McClain, 2011), the results of which can be seen in Appendix B, this research utilized the tool as it operates in conjunction with the web spider. Essentially, the spider sits on top of the STANLEY program, and once the spider results are returned, the tool calculates scores and statistical information for use in analysis. The set up and execution of the tool was accomplished at Sandia's laboratory in Albuquerque, NM based on the research questions addressed in this thesis. Given the question at hand and the list of seed URLs and documents, Mr. McClain instantiated and executed the tool and forwarded the results for analysis.

### **4. Establishing the Document Corpus**

The document corpus used to determine a correlation between Mid-East turmoil and U.S. oil prices was formed from the results of the spider and tested for similarity against the seed documents. To simplify the analysis, and ease the manner of comparing the oil price data with the document scores, a date extraction method was utilized to glean the dates of the articles from the various websites. The period of interest was from January to June 2011. While the extraction method was not able to provide accurate

dating for all the results, it was able to provide moderately sufficient results for analysis. For the results where dates were returned, the date extraction was only accurate to the month and year.

## **5. Scoring the Documents**

Using the cosine similarity function discussed in II.D, the results from the spider were scored against the seed documents (Pant, Srinivasan, & Menczer, 2011). The website pages were processed individually to extract dates and compare the results to a statistical model of the source documents (Bauer et al., 2005). These scores were initially calculated for every returned document. The date extraction algorithm used two approaches to extracting dates. It first checked for server assigned dates, such as “last modified” or “server expiration” dates. If no dates were found (and generally there were not), entity extraction was used to extract dates from the unstructured text of the page. After date extraction reduced the number of results and the documents were binned by month, the scores calculated were the monthly average (mean), mode, median, minimum, and maximum values for the documents returned for a given month. It was these values, particularly the mean and median, which were used in the analysis of the tool’s effectiveness. Confidence was also calculated and tracked based on the percentage of the total number of documents obtained for a given month. These calculations, returned from STANLEY in .csv format in Excel, were organized by month and transferred to JMP Pro 9 for analysis and comparison with the oil price data.

## **6. Methods of Analysis**

The primary mode of analysis for STANLEY used JMP Pro 9 statistical software to determine the correlation coefficient between the real-world oil prices and the similarity scores calculated by the tool. These scores were entered into one Excel spreadsheet where they were organized into individual website sheets and a sheet containing the combined and average scores for all websites. They were then compared with the oil price data. Correlation matrices and pair-wise correlation factors were determined from the results, and simple linear regressions were conducted to compute the fit of the scores.

## **IV. ANALYSIS**

For each tool, the results of the applications needed to be analyzed in order to provide some manner of validation for the tool and justify its use as a potential agent instantiator for social simulations. Unless otherwise noted, the tool for analyzing and comparing the data was JMP Pro 9. This software, developed by the SAS Institute, is a potent statistical and analytical tool that enables the user to interact with the data in powerful ways. It also allows for graphical representation of the data in manners that spreadsheets cannot (JMP, 2011). The methods of investigation, overall, were relatively straight forward. Simple statistical analysis was used to determine the accuracy of the survey tool's CG results for national satisfaction and correlation for the STANLEY tool's test with oil prices and Middle East turmoil

### **A. SURVEY DATA DEVELOPMENT TOOL**

The main purpose of this portion of the thesis was to determine if the Survey Data Development Tool could produce effective case files that could be used to create agents for use in populating social simulations. In the case of this research, the CG model was populated using agents developed from case files generated from surveys conducted by the Pew Research Center.

The first area that needed to be verified was to determine if the case files generated by the tool could in fact instantiate the tool and initialize the Bayesian networks. This was done by creating five ~~hand-made~~" case files based on the data in the actual case files. This smaller batch of case files was used to generate agents within the CG model that ~~resided~~" in each of four regions. One of each stereotype was placed in each region (the regions corresponded to the regions defined in the survey and explained in III.A.9.a.) resulting in a 20 agent system. The test was run for one replication constituting a period of 90 ~~days~~." The purpose of this mini-test of the model was to check that the agents were communicating with one another, that satisfaction was changing over time, and that sentiments were affected as stimuli were injected at various

times in the model run. These stimuli were changes to any of the oil price, unemployment, or consumer economic confidence indexes.

Figure 31 consists of a generic version of Figures 24 and 25. It is a graphical depiction of what happens to an individual agent within CG when an indicator is changed at a given time in the model run. Each of the regions represents an agent's belief network at a time in the run where an indicator change occurs. (In the full model runs, these times correspond to actual survey dates.) The gray arrows represent a stimulus injection at the particular time changing one or more of the indicators (noted in the arrow). –IssueStance” signifies the issue in question in the network (national satisfaction). –Belief” represents the agents’ beliefs and responses to survey questions.

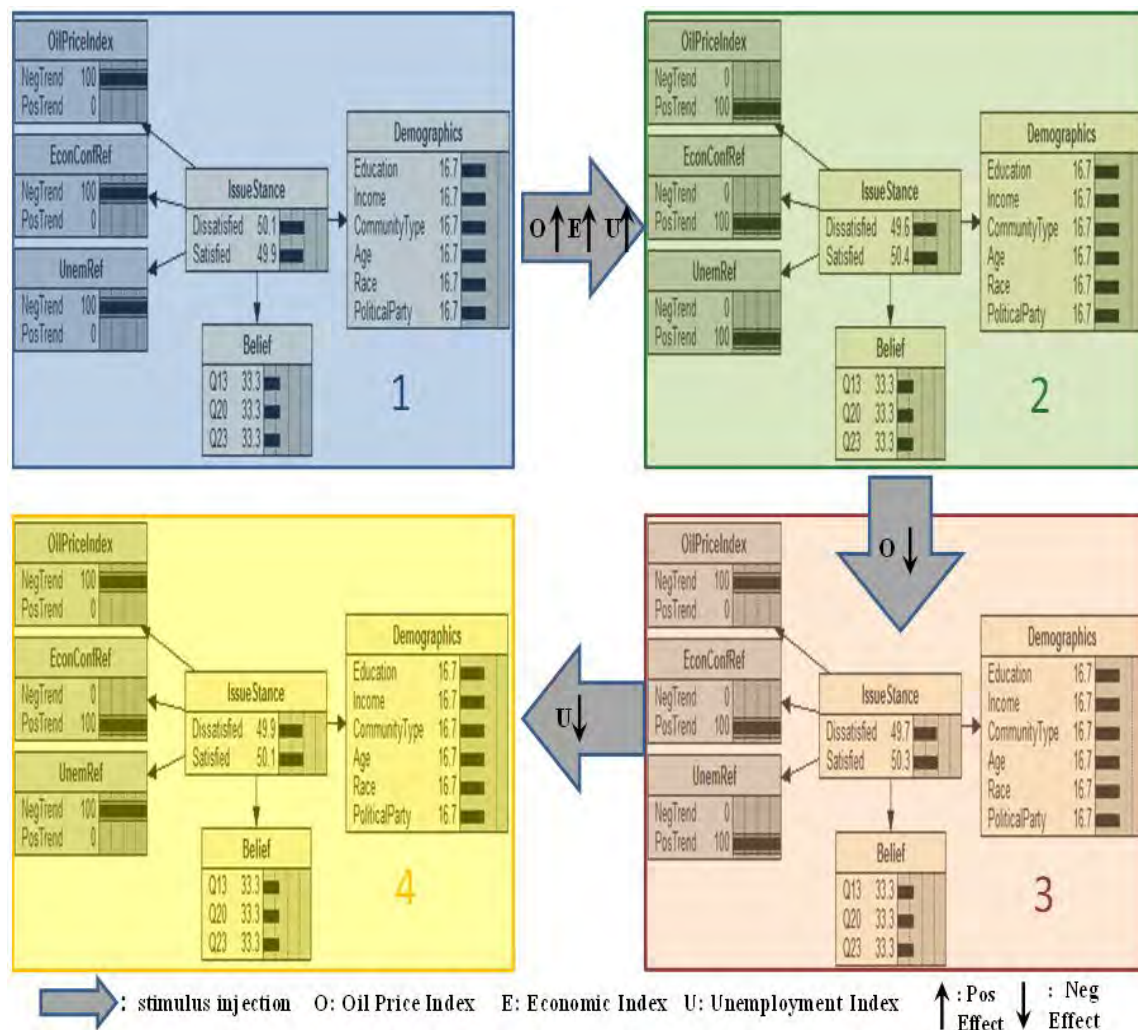


Figure 31. Illustration of BN changes due to indicator shifts (best viewed in color)

As an indicator changes (i.e., from  $\neg$ NegTrend” to  $\neg$ PosTrend”), its value in the BN becomes 100% for whichever trend is perceived. The levels (probabilities) for satisfaction and dissatisfaction within  $\neg$ IssueStance” increase or decrease accordingly (based on Bayes’ Rule) and are reported in the CG output. This is illustrated in the networks by the regions in the figure. Region 1 represents initial conditions in the model, and each subsequent region shows the changes to the probabilities affected by an indicator change. In very simple terms, after the issue stance is updated, the networks essentially “reset” at the new probabilities (a new initial condition) and are ready for the next change in an indicator. Over the course of the model run, the affects of the changing indicators on the agents’ issue stances are cumulative, resulting in a continuum of change in satisfaction over time. It should be noted that the changes seen in Figure 31 are simulated and do not represent the actual probabilities in the study.

The tests conducted with the 20 agents proved that the agents were communicating and the CG model was responding predictably (expected output) based on the provided input. In other words, the agents’ satisfaction was increasing and decreasing over time. They also showed that the established Bayesian networks were set up correctly and appropriately feeding the model. Figure 32 shows the preliminary results from this small run. The vertical reference lines indicate where stimuli in the form of an index change occurred. The three lines plotted in the chart represent a small example of the numerous changes to indicators throughout the simulation runs. Note that there are shifts in agent satisfaction at these locations which correspond to the direction of the indicator changes. For example, at time 36, the oil price index decreased (i.e., there was a shift from a  $\neg$ PosTrend” to a  $\neg$ NegTrend”). Agents responded by decreasing satisfaction. This was an encouraging result, as it showed that the model was responding to the environment. With these results, full model tests were set up and conducted using the complete set of case files created by the tool for both the training and full survey data sets.

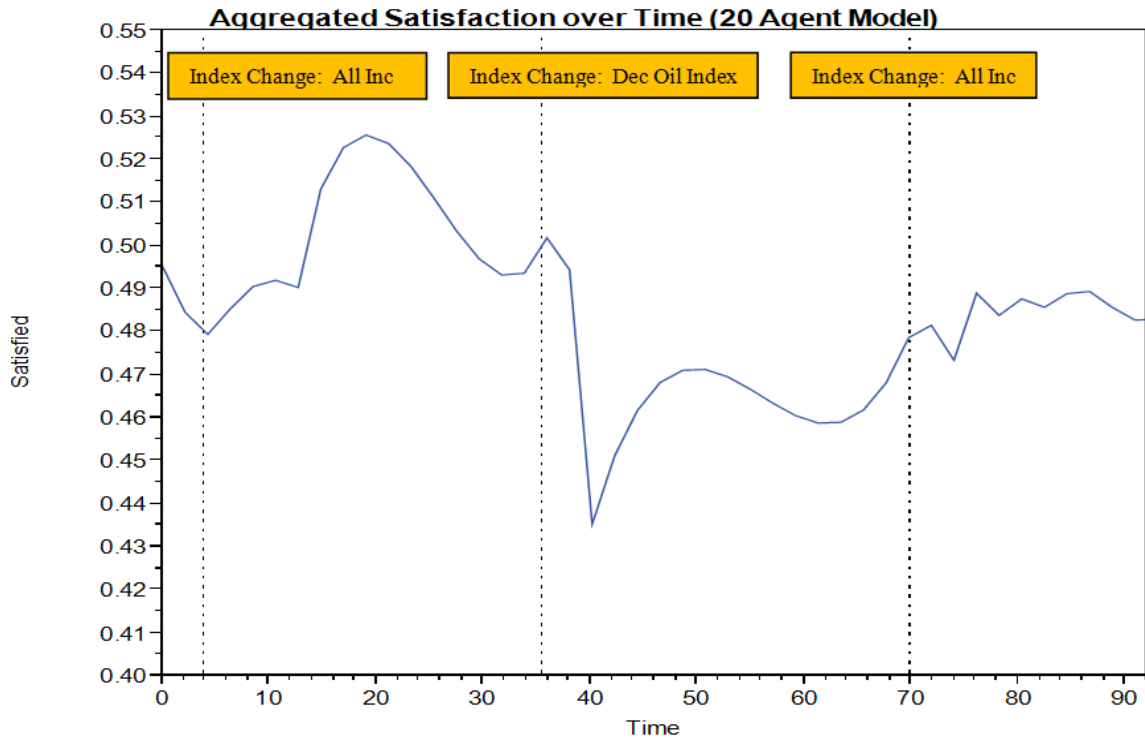


Figure 32. Twenty agent test results with reference lines showing index changes

Also of note from the 20 agent survey was the discovery of which of the six demographics had the most significant impact on satisfaction or dissatisfaction within the agents. While not critical to this research, it demonstrates an important aspect of CG, in that it can allow an analyst to determine which demographics are most relevant to survey results. It can also certainly be applicable to future analysis, and thus it is included here. After some investigation with all the combinations of satisfaction among demographics, *Age*, *Race*, and *Income* were the most relevant to the satisfaction level, from a purely demographic standpoint. Figure 33 shows graphically these relationships. For an agent to be “satisfied,” it had to have a satisfaction score, as returned in the CG output as probability of satisfaction, greater than 0.50 (50%). The shaded regions in the lower three sections of the figure indicate the number of “dissatisfied” agents in a particular demographic set (as a fraction of total agents in that demographic set). The frequency counts indicate the number and fraction of agents represented by the shaded regions.

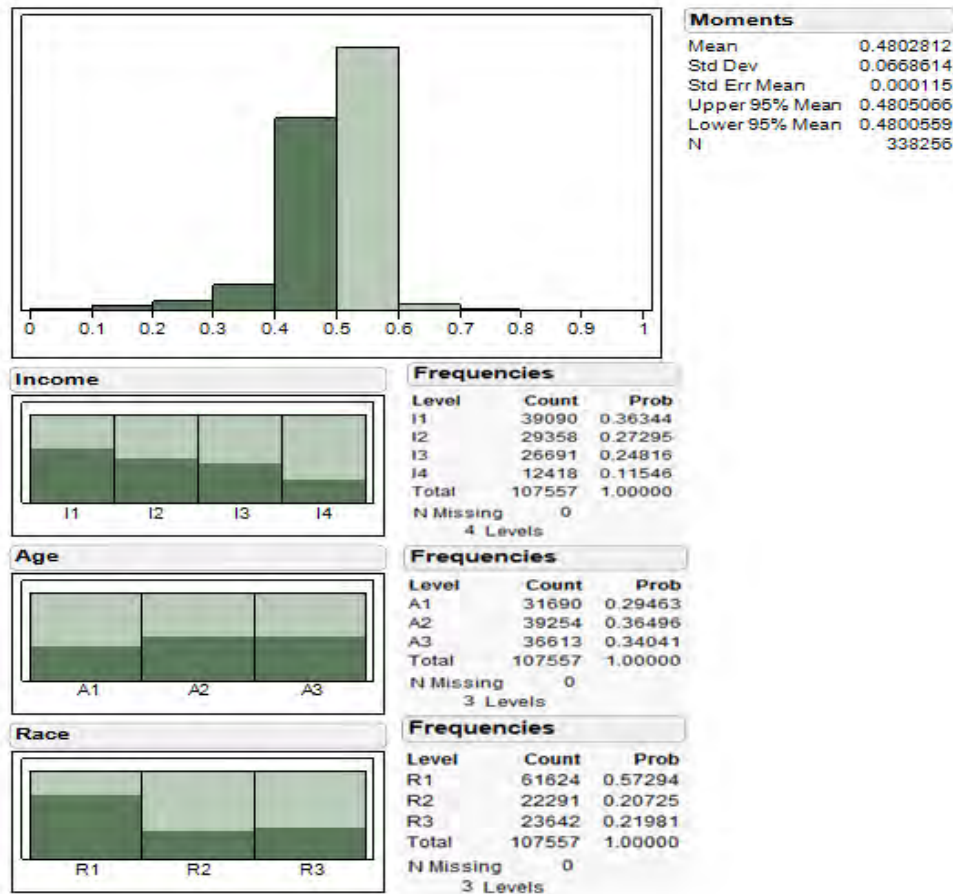


Figure 33. Relationship of most relevant demographics to satisfaction

Three scenarios were developed for the full model run. Each was run for five replications of 260 days. Historical knowledge of the CG model, coupled with results from recent applications, yield that after more than five replications, the long term results tend to be very similar, so for the sake of time, the number of replications was limited to five. Figure 34 shows this phenomenon. Note how all the replications are nearly identical.

The first scenario developed, hence referred to as CG1, utilized case files generated from the training set of survey data. The second and third scenarios (CG2 and CG3, respectively), contained the full set of survey data. CG2 used the original case files generated by the tool (no minimum number of rows), and CG3 used case files with 10 rows minimum of data (additional rows created by sampling with replacement). The CG

results, unless stated as ~~normalized~~,” are percentages of satisfaction for each agent, on a scale of zero to one, where one is completely satisfied. This is how the satisfaction results from the CG model are measured. When a satisfaction score is termed ~~normalized~~” or it is compared directly with the Pew data, it represents the percentage of the population (agents) which responded ~~satisfied~~” to the question of national satisfaction, since this is how the Pew results were represented in the final reports generated by the Pew Research Center.

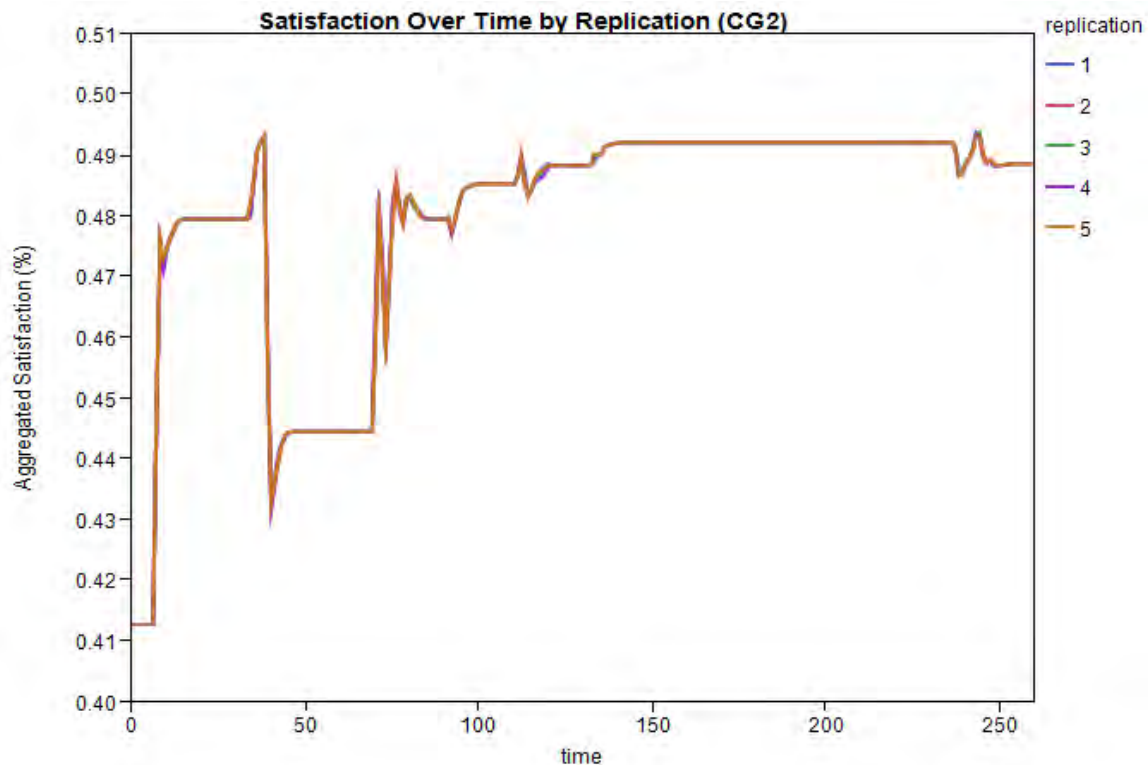


Figure 34. Satisfaction over time for CG2 for all replications (best viewed in color)

As a point of interest, Figure 35 plots the satisfaction over time for each agent in CG2. This is included to show that each agent does in fact react differently to the stimuli throughout the simulation run. Note the periods of flat lines. These indicate the time in between stimuli injection points within the scenario run and show that the model is not creating any additional behaviors within the agents except at those times specifically designated in the input tables. It also proves that the assumption of index change alone being sufficient to alter satisfaction was initially valid. (There was a seven-day

memory” period for each agent, so there were actions up to seven days after a designated injection, but nothing occurred after that until the next specified time.)

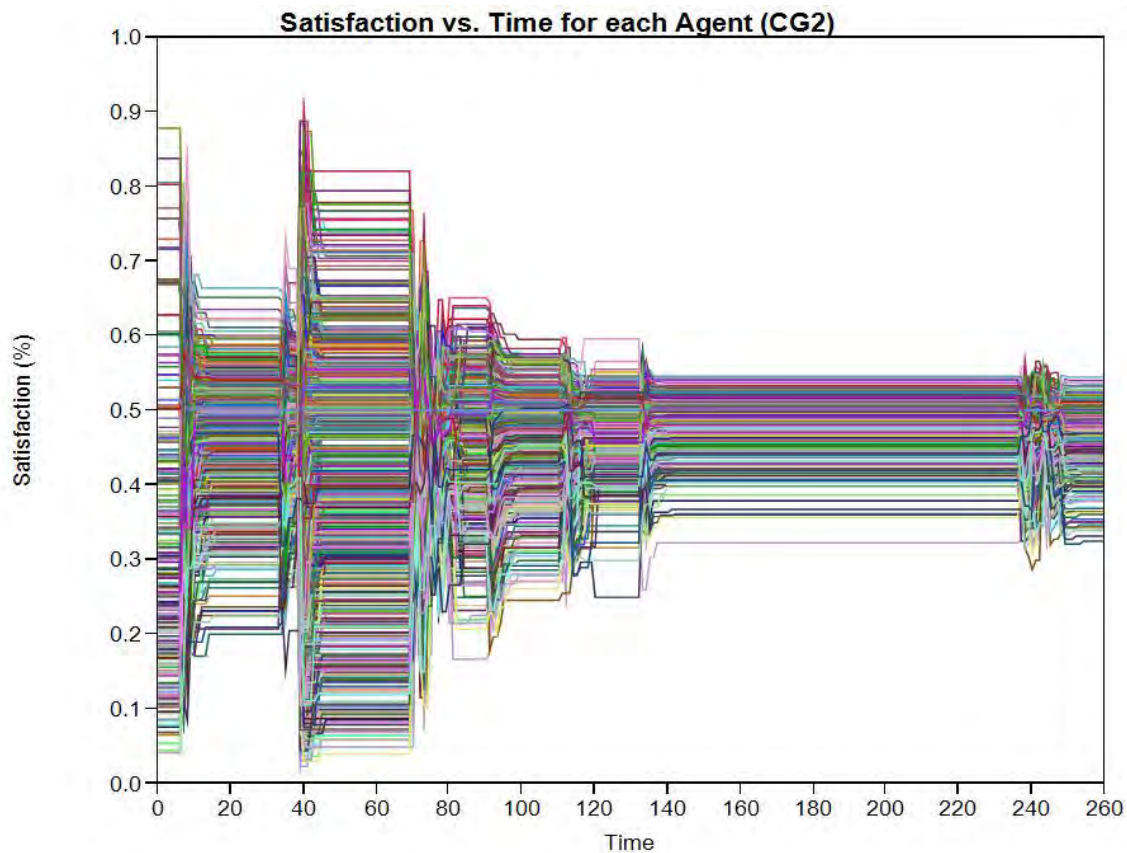


Figure 35. Plot of satisfaction over time for each agent (CG2) (best viewed in color)

CG1 tested the model’s behavior after a three ~~month~~” training period and was compared to the Pew results. Figure 36 shows the trained model satisfaction over time, and Figure 37 shows the model’s results and the actual Pew results over the time period of the study, with satisfaction percentages normalized representing agents’ and respondents’ satisfaction. Figure 38 shows how CG2 and CG3 compare over time. The significant difference between the two was the inclusion of a minimum row per case file requirement for CG3. The overall similarity in the lines indicates that in this case, the minimum row requirement did not significantly affect the outcome. JMP analysis comparing the differences of the means of CG2 and CG3 over time (using nonparametric tests for means) confirms the observation. Figure 39 shows the results of this test.

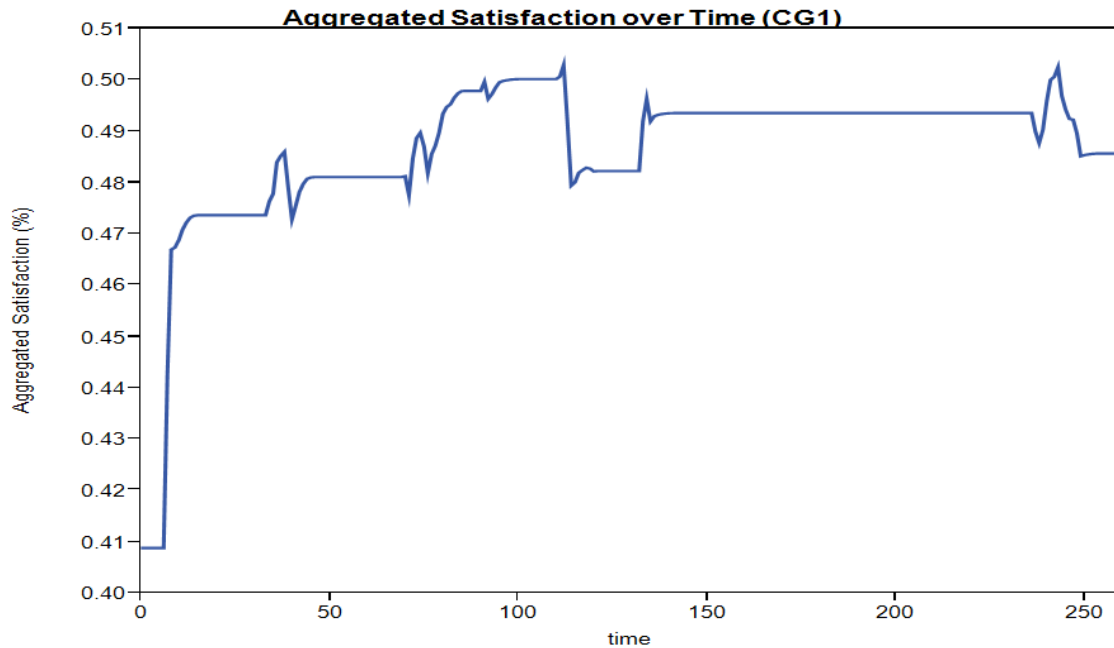


Figure 36. Trained CG model satisfaction over time

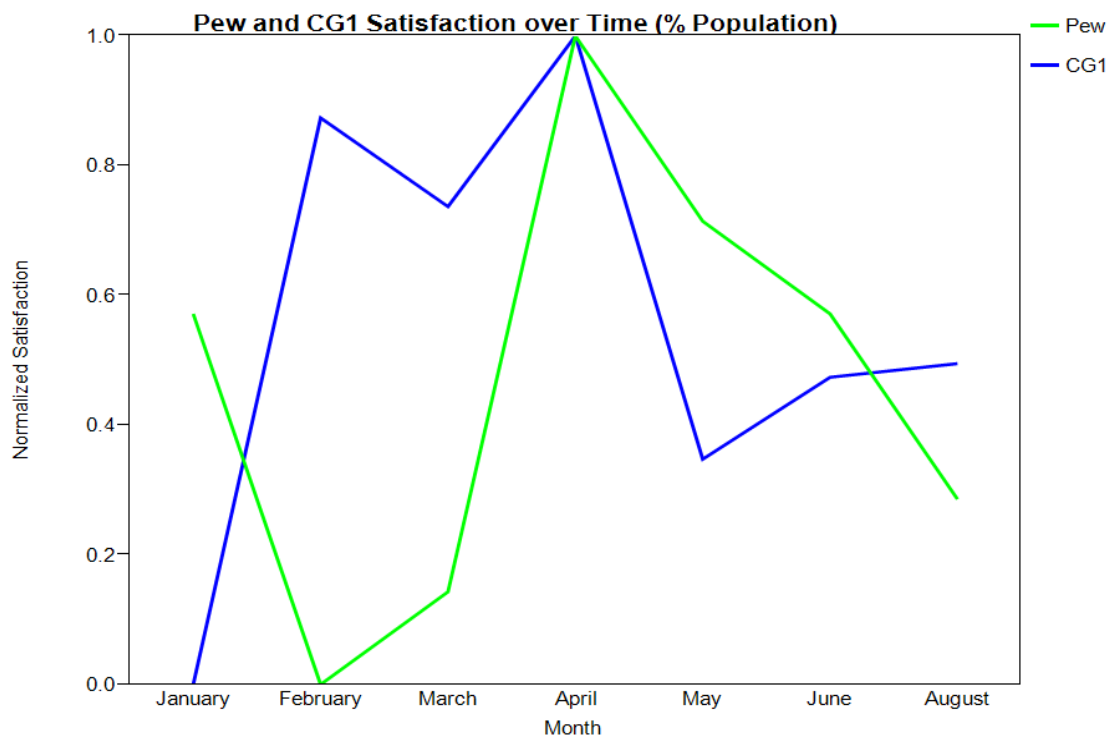


Figure 37. Trained CG model vs. Pew results over time (best viewed in color)

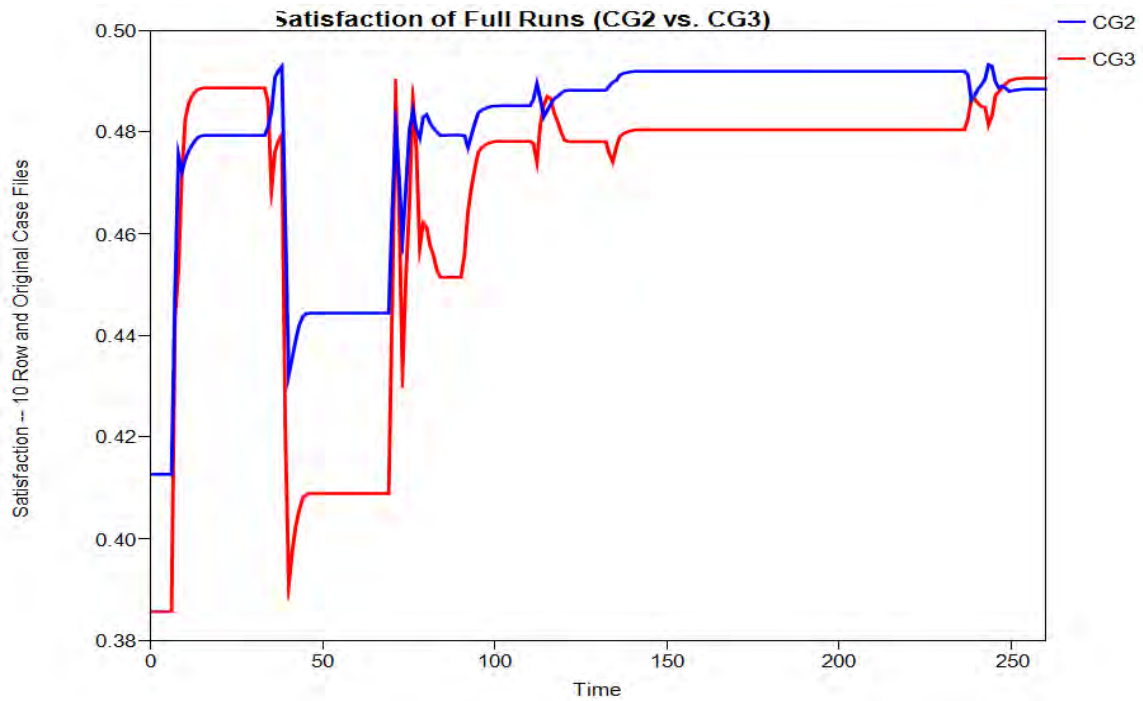


Figure 38. Aggregated satisfaction over time for full model run (CG2 & CG3) (best viewed in color)

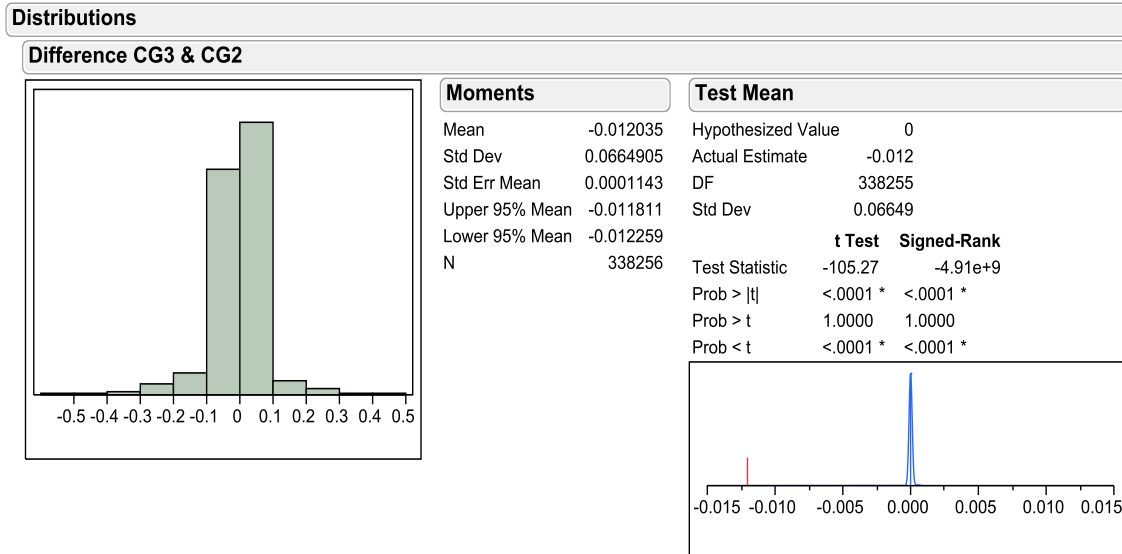


Figure 39. Statistical analysis on significant differences between CG2 and CG3

CG3 was then compared against CG1 and the Pew data to check the models' long term behavior. From these full tests, plots were generated of satisfaction over time with which to compare the actual Pew survey results. This comparison was done visually and showed that the satisfaction overtime was somewhat similar in shape for the sets of results. See Figure 40. Of particular note, the areas of interest are where the different lines trend in the same direction following stimuli injection. CG output matched the Pew data in direction of sentiment towards satisfaction as the model progressed, but at the start of the simulation, there were significant variations between the two results. To bring the sets of results to the same scale, the percentages for the Pew satisfaction and CG results were then normalized and plotted again. This allowed a much closer look at how the sets of data trend over time. It was quite apparent that there were some issues affecting the models' output during the first 60 days or so of the run, as they diverge drastically from the real world results (see Figure 41).

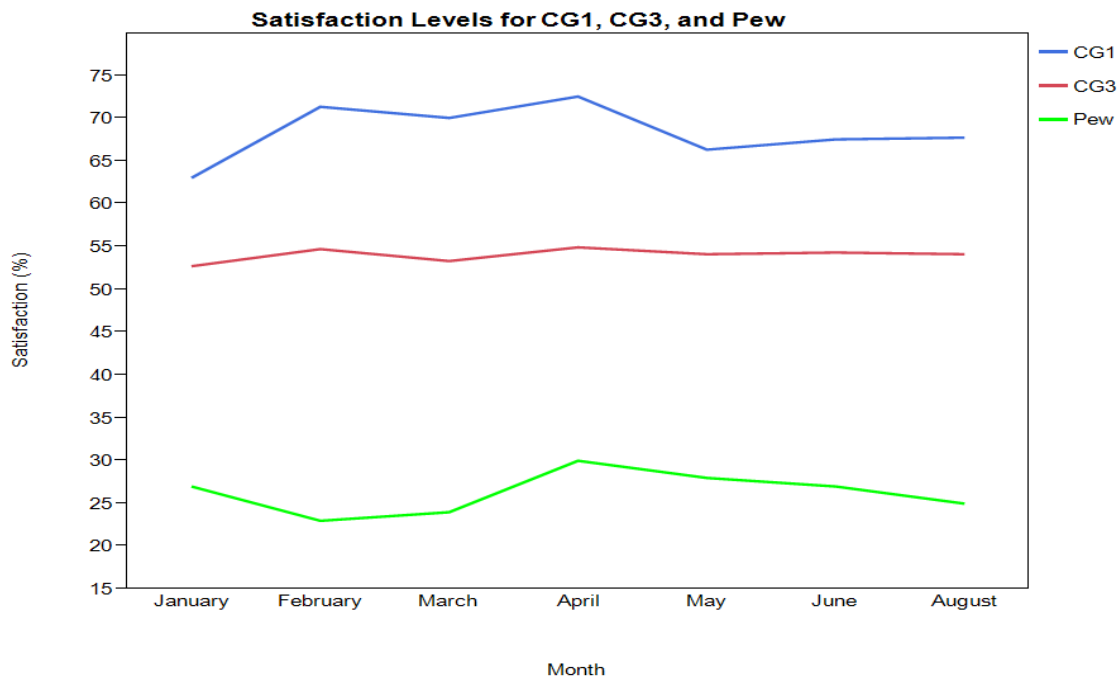


Figure 40. Respondent/Agent Satisfaction percentages for CG1, CG3, and Pew Data sets (best viewed in color)

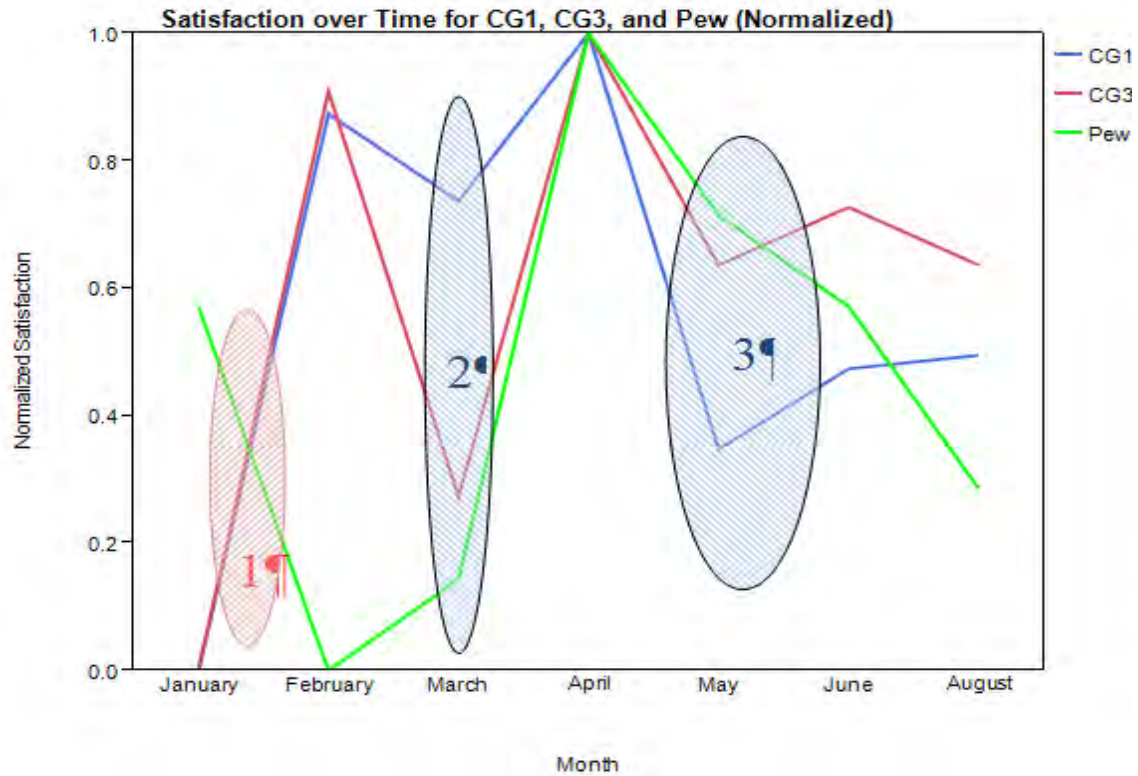


Figure 41. Aggregated satisfaction percentages (normalized) (best viewed in color)

## B. DISCUSSION OF ANALYSIS

As mentioned earlier, the primary means of comparing the results involved examining the trends over time of the CG output to Pew data. When the plots are analyzed, it is obvious that the CG and Pew results initialize at different levels of satisfaction. They then diverge until about March (see Figures 37, 40, and 41). From March on, the results exhibit more similar behavior until the end of the simulation run, where they tend to diverge yet again. The numbered regions in Figure 41 highlight the trend analysis. Region 1 shows where the results diverged. There are several possible causes for this. First, there may have been some warm up time needed for the agents to begin behaving predictably. Second, the potential warm up may have led to some artifacts generated that CG handled in a particular way. These artifacts, present until the model has warmed up, could cause unpredictable behavior. Third, and perhaps most applicable, no real world events, short of the three indexes discussed, were applied to CG. One of the underlying assumptions was that changes in the three indexes would be

sufficient to shift an agent's satisfaction. February of 2010 was a time of much political debate in the country, as Congress debated President Obama's health care legislation ("2010 Current Events," 2010). This was not accounted for in the model, and lends that the assumption that the three indexes alone would cause appropriate changes in satisfaction could be insufficient. Regions 2 and 3 show where the trends follow in the direction of each other. There is some divergence in Region 3, but that may be due, in part, to the British Petroleum oil spill that occurred in the Gulf of Mexico during this time period. Overall, these are encouraging and expected results.

Despite the stated divergences in the plots and the possible causes for these phenomena, the CG model results show and the analysis confirms that the Survey Data Development Tool does in fact create case files that can be used to populate the CG model, and social simulations in general. Additionally, preliminary inspection indicates that the minimum row requirement for each case file does not significantly alter the output of the model. It should be noted that there were some limitations to the survey data that might have affected this result. Not all questions included in the Pew survey data used for this thesis had responses, so the sampling with replacement done by the tool on these case files may have resulted in redundant rows in the case file. In fact, a quick inspection of several case files revealed this. It appears that the minimum row requirement and the sample with replacement method might be more effective given a more diverse response set. This is an area for future examination.

### **C. STANLEY**

The text analysis tool from Sandia was tested to see if it could aid in determining the existence of a correlation between the 2011 "Arab Spring," the name given to the turmoil and political change present in Middle Eastern and North African nations in early 2011, and the rise in U.S. oil prices during that same time. The tool automatically scored the results, based on the cosine similarity function, against the seed set. These scores, which included the mean, median, mode, maximum, minimum, and range of scores of a website's returns, were normalized (using equation 1.4) and compared to normalized oil prices across the same period of time. An assumption regarding these scores stated that

scores of zero (0) indicated an inability to score a document, thus those scores were removed from the results. As mentioned previously, a current limitation of the tool is that the date extracted by the spider is only reliable to the month and year. This unfortunate limitation resulted in only six data points for analysis, although there were five websites returned which added to the sample number.

$$1 - \left( \frac{Max - Val}{Max - Min} \right) \quad (1.4)$$

Where *Max* is the maximum value in a given column, *Min* is the minimum value in the column, and *Val* is the value of interest.

Once the data and scores were entered into JMP Pro 9, several methods were used to explore the results. Initially, simple graphs were made in which the normalized oil prices were plotted with the mean document scores against time (see Figure 42). This figure is busy, and while it does not show any clear correlation between the oil prices and the document similarity scores, it does show that the scores for [www.forbes.com](http://www.forbes.com) and the mean scores for all sites (consisting of averages for all the scores of all five sites and henceforth known as “All”) more closely follow the trends of oil prices than the other sites. Figure 43 shows this relationship more clearly. ([www.thestreet.com](http://www.thestreet.com) only returned data for one month, so it was excluded from the results.) Since [www.forbes.com](http://www.forbes.com) and the scores for “All” appeared to have a stronger visual correlation than the other sites, they were the two references used for comparison with the oil prices. Additionally, mean and median visually showed the strongest correlation (see Figure 44) and were used in the analysis.

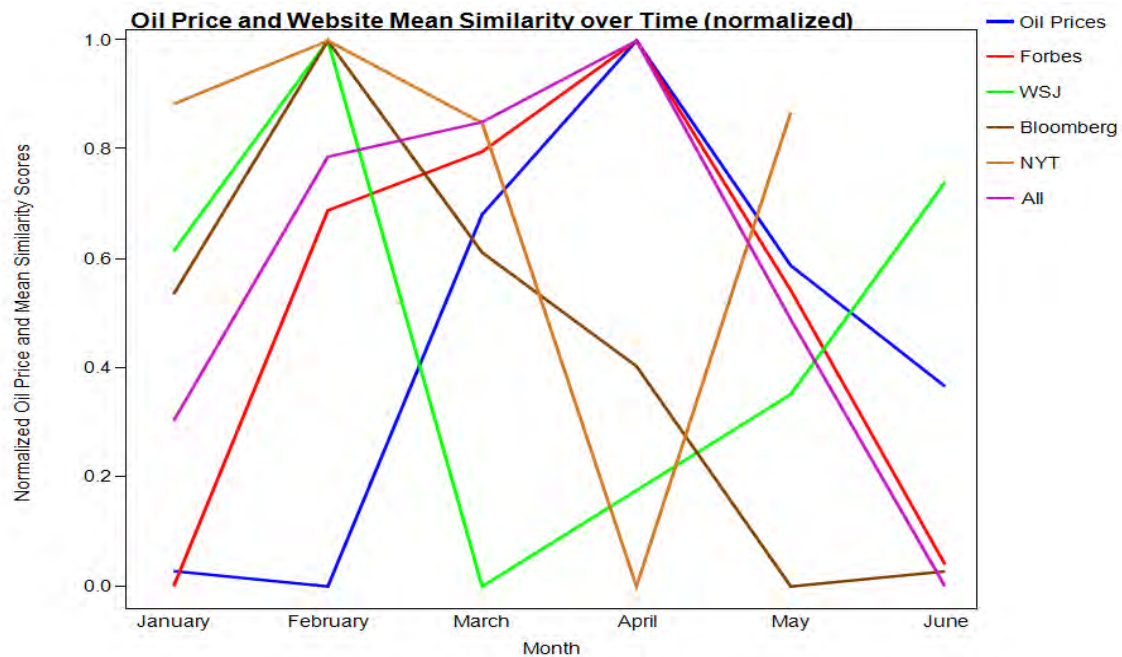


Figure 42. Oil prices and spider scores over time (normalized) (best viewed in color)

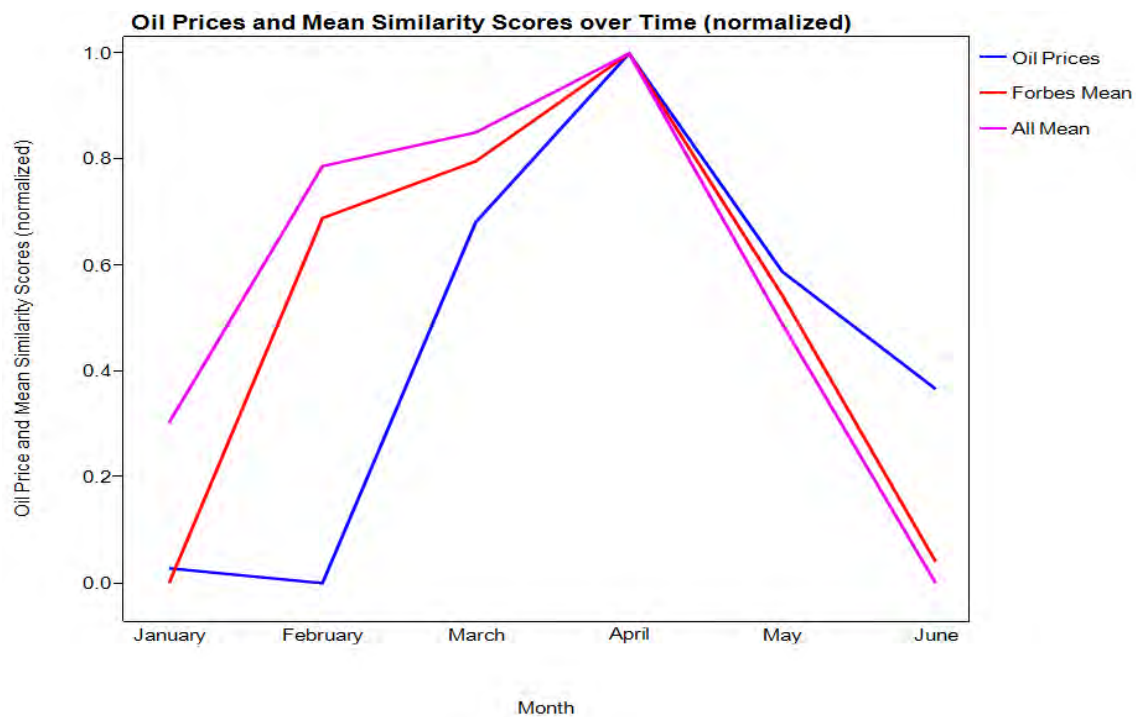


Figure 43. Oil prices, forbes.com and all sites score (means, normalized) (best viewed in color)

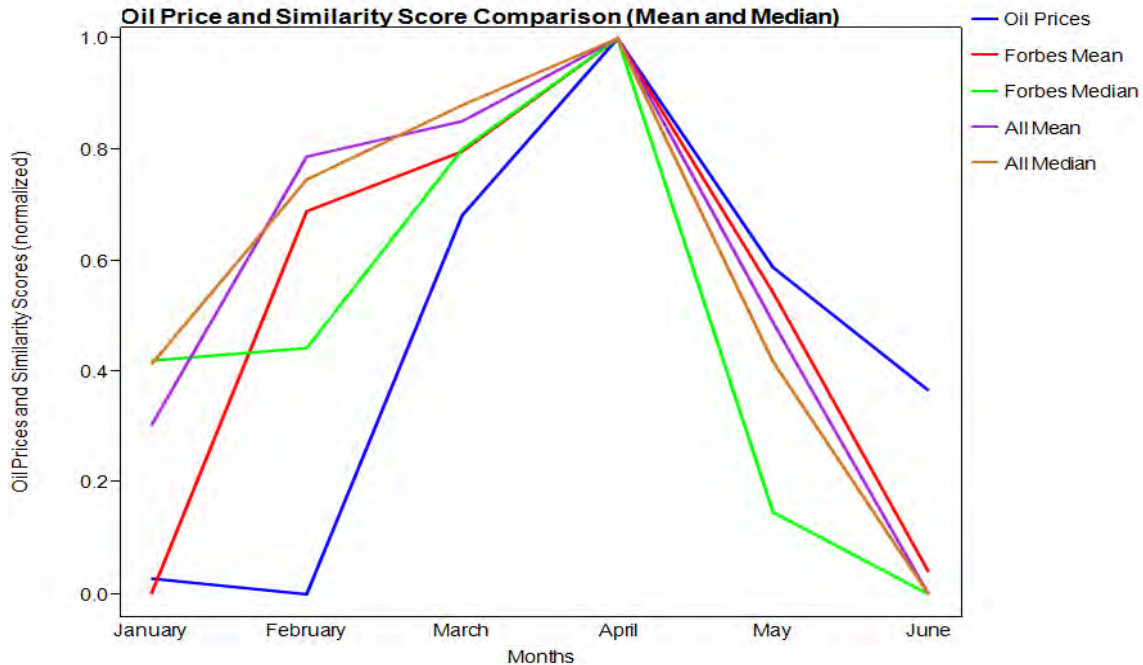


Figure 44. Mean and median scores with oil prices over time (normalized) (best viewed in color)

The plots do a good job of visualizing the correlation that exists between the narratives about the “Arab Spring” and the increase in U.S oil prices during that time. Even though only five sites were examined, it can be seen that the similarity scores and oil prices peak at the same time, in April. Furthermore, they all tend to trend in the same direction over the period of interest, except for the first two months (January and February). It is unknown what led to this small divergence. It is more than likely due to the fact that the uprising and turmoil in the Middle East and North Africa had not yet really spread and become more broadly reported on, nor were the effects on oil prices realized yet.

After the plots were created, a multivariate analysis was conducted. Multivariate scatterplots were generated consisting of the oil prices and document scores (mean and median, normalized) from [www.forbes.com](http://www.forbes.com) and the combined results from all sites (Figure 45). These results show a correlation range of 0.446 to 0.644 for the scores of interest. Pair-wise correlation tests were conducted (Figure 46) as well as simple linear regression analysis to look at oil prices versus similarity scores (the means in this case, as

they had higher correlations). The results are shown in Figures 47 and 48. The fit of the lines generated R-squared values of 0.415 (Forbes) and 0.229 (All). These indicate that between 22.9% and 41.5% of the variation in oil prices is explained by the sentiment expressed by documents gleaned from the sites explored in this research. (Compared with the results shown in Appendix E, these are significantly improved and lend credibility to the assumption that led to the disregarding of zero value scores.)

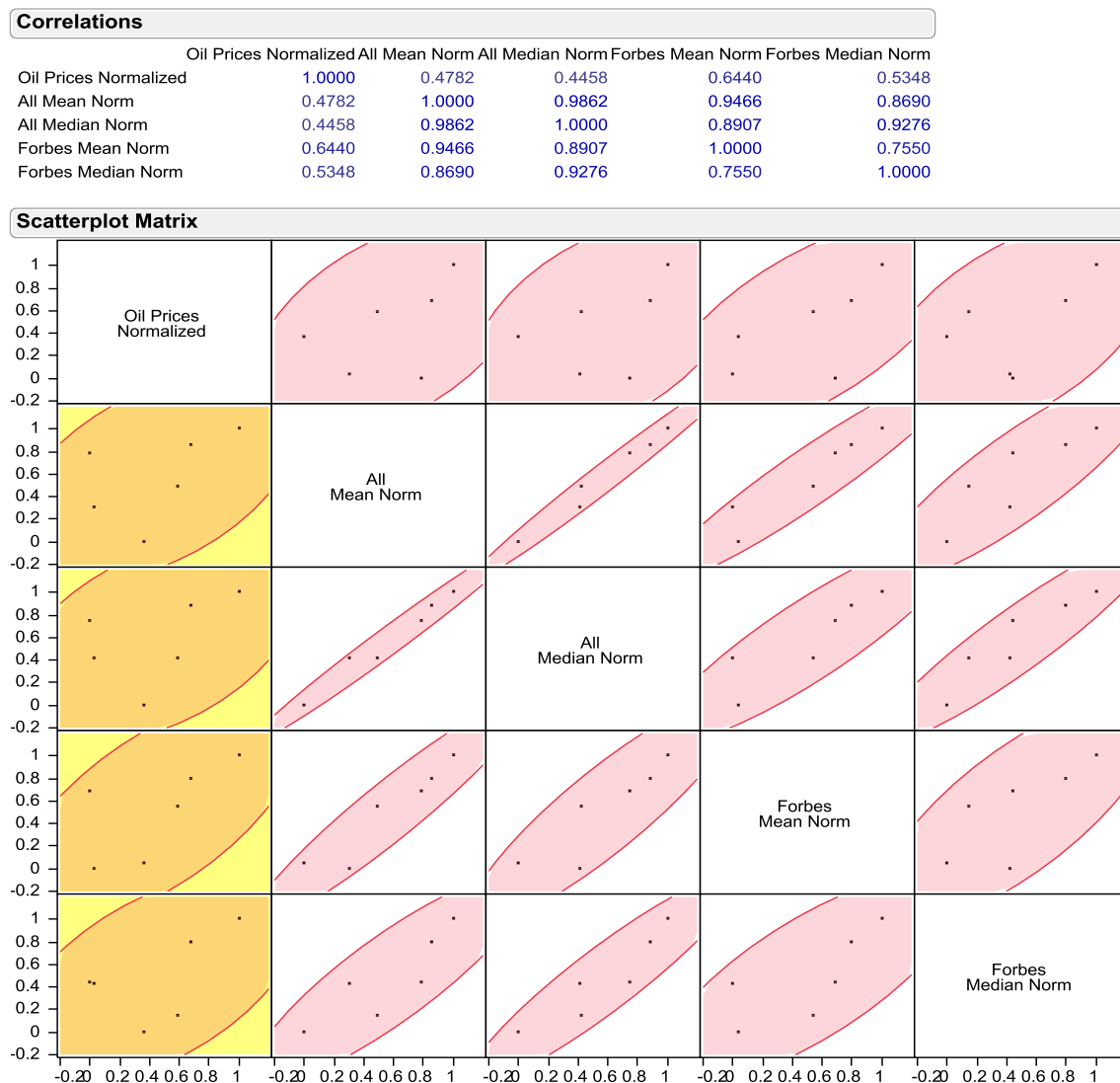


Figure 45. Scatterplot matrix showing correlations of oil price to similarity scores

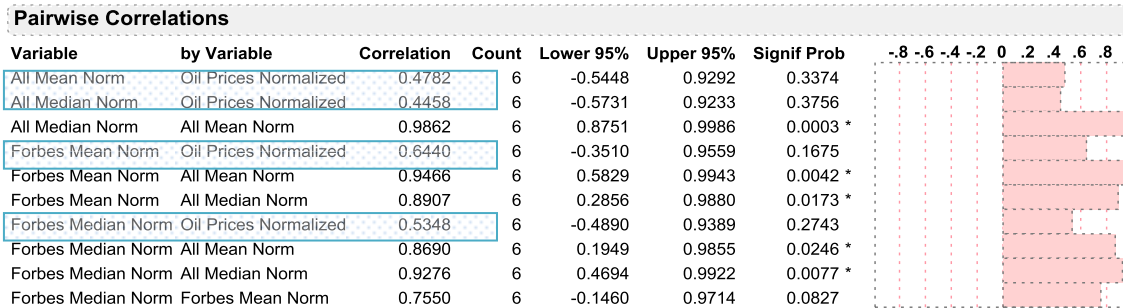


Figure 46. Pair-wise correlations among oil prices and similarity scores

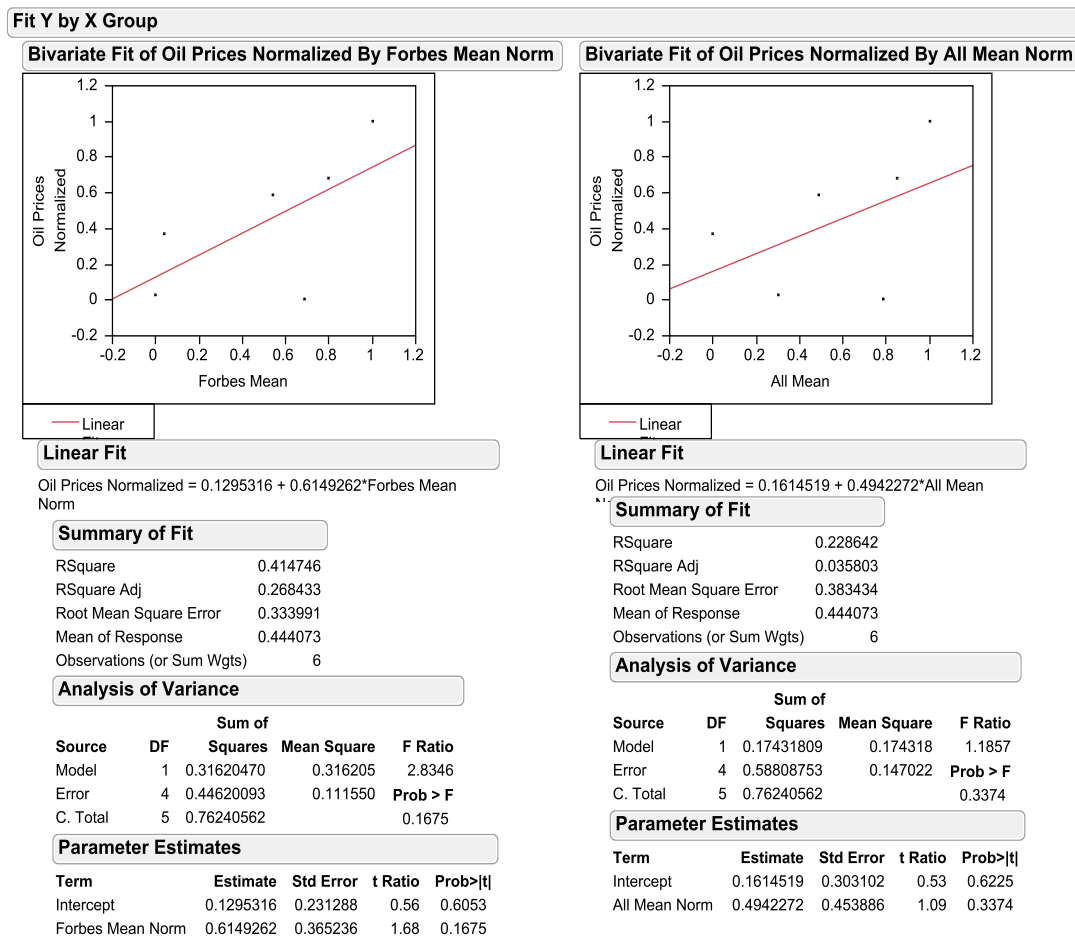


Figure 47. Regression analysis for oil prices with mean similarity scores (Forbes and All)

Response Oil Prices Normalized				
Summary of Fit				
RSquare		0.58115		
RSquare Adj		0.301917		
Root Mean Square Error		0.326258		
Mean of Response		0.444073		
Observations (or Sum Wgts)		6		
Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	0.44307218	0.221536	2.0812
Error	3	0.31933344	0.106444	Prob > F
C. Total	5	0.76240562		0.2711
Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	-2.914278	1.95278	-1.49	0.2324
All Mean Norm	-1.308177	1.198265	-1.09	0.3548
Forbes Mean	35.767468	22.50983	1.59	0.2103

Figure 48. Summary of fit results for oil prices and Forbes and –All” mean similarity scores (combined)

#### D. DISCUSSION OF ANALYSIS

While the graphical and analytical results indicate some correlation between the scored results and oil prices (0.644 for Forbes mean and 0.478 for –All” mean), an insufficient number of data points makes it impractical to definitively state whether these results would hold true for a larger sample size, or if this was just a random occurrence given the particular conditions experienced in this study. While it cannot be said whether or not the tool is successful analyzing and forecasting sentiment, the results gathered here, especially given the assumption that zero score documents can be removed, indicate that the tool has potential, particularly if it operates under certain caveats. R-squared values from the summary of fit for the individual indexes (Forbes and –All,” Figure 47) and the combination of the indexes (Figure 48), while too low to suggest an obvious relationship between the oil prices and sentiment, are high enough to prompt continued investigation into better ways to maximize the tool’s capabilities given its current limitations (limitations that can hopefully be remedied).

Another point of discussion involves the variability in the similarity scores. The scores varied dramatically for each site used in this study. For example,

[www.forbes.com](http://www.forbes.com) had scores ranging from near zero (zero scores were removed) to 0.252 while [www.bloomberg.com](http://www.bloomberg.com) had a maximum score of 0.292 and [www.nyt.com](http://www.nyt.com) had a maximum of 0.321. Figure 49 plots the maximum similarity scores for four of the sites, and Figure 50 plots the means of the scores. These plots show the wide range and disparity of returned scores for the sites across the period of interest. Still, the similarity scores were quite low overall. The maximum value returned among the websites crawled was 0.321. McKaughan, et al. (2011), in their previous experimentation with STANLEY, routinely saw similarity scores over 0.6, with some as high as 0.9. This may reflect on the lack of strength and orthogonality of the classifiers for this study, as well as the less focused results returned by the spider. Whether these score levels are particular to the time period investigated here or if they would be similar given the seed documents and a larger time period is for future research. These issues with the tool and the data it returned significantly impacted the study and the ability to draw any conclusive opinions about the effectiveness of the tool.

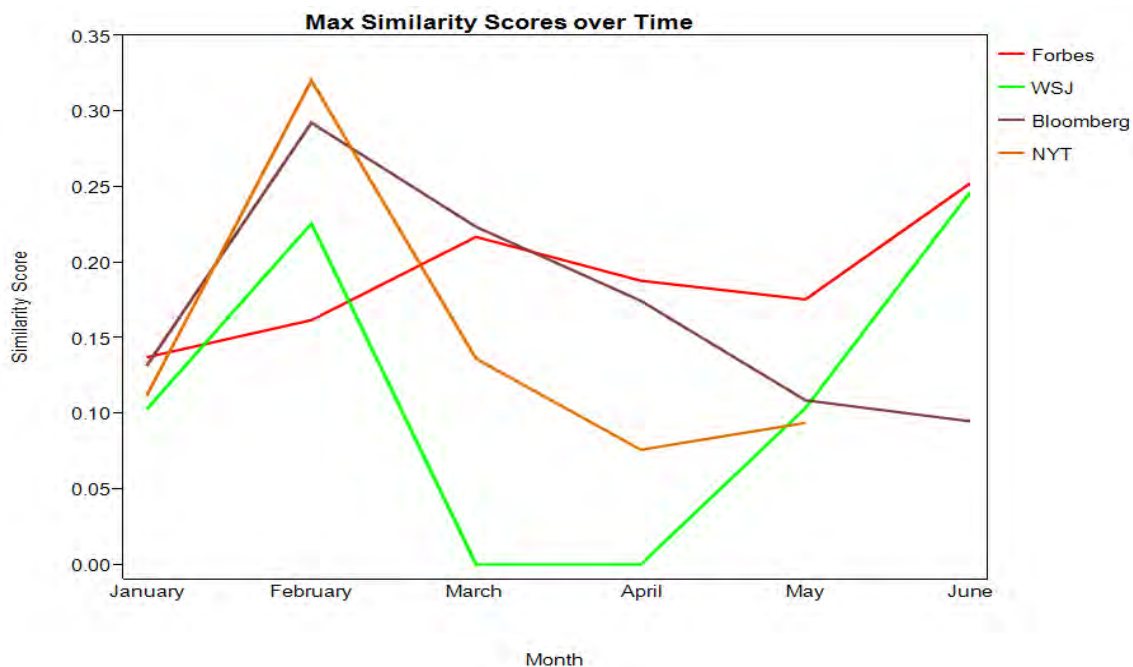


Figure 49. Maximum similarity scores of websites over time (best viewed in color)

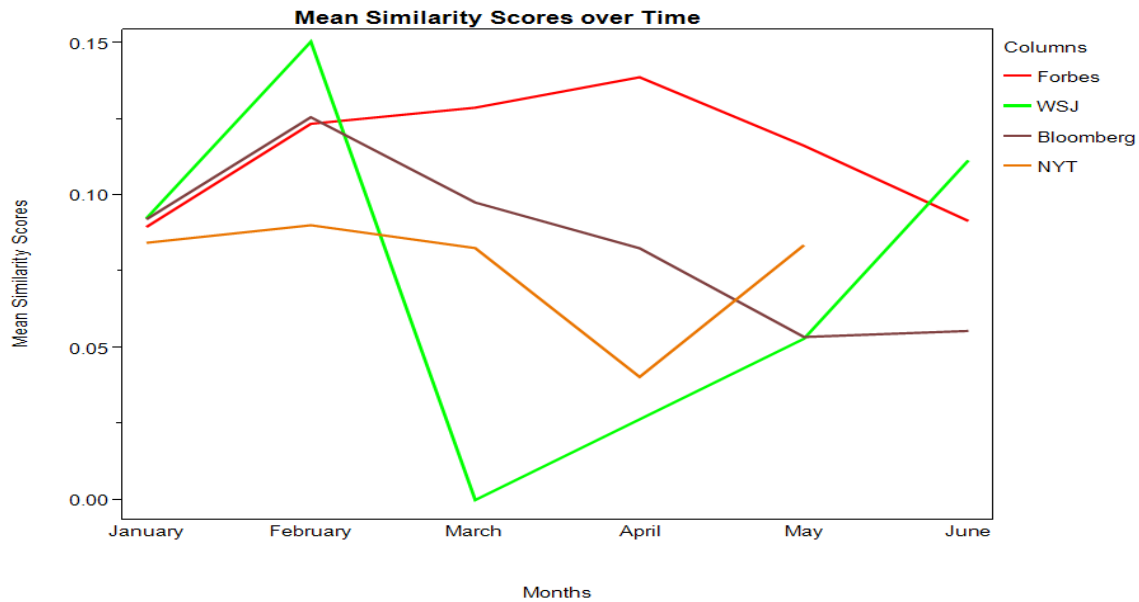


Figure 50. Mean scores of websites over time (best viewed in color)

However, not all is lost. A low sample size resulting from the tool's inability to more specifically determine dates was the greatest single inhibitor to the success of the research. Had the tool been more capable in its date extraction, the results may have been significantly different. The very limited results shown by the correlation matrices and regression models lead one to believe that with more data points, a stronger correlation and higher R-squared could have been discovered. This is certainly an area for future research.

## **V. CONCLUSIONS AND RECOMMENDATIONS**

As mentioned in the Introduction to this thesis, counterinsurgency operations and irregular warfare have become a major facet of U.S. military and Department of Defense engagements. While the nation has not neglected major combat operations, as threats are emerging in various parts of the globe, the currently pressing matters involve localized conflicts against ideology and extremism. The keys to victory are in winning the hearts and minds of the local populations where these extremists exist.

The Cultural Geography model, like other social simulations, attempts to model the responses, behaviors, and actions of a local population, given various operations or accomplishments completed upon said population. By understanding how a society might respond, commanders and planners can shape their actions to maximize the positive impacts on local groups while achieving strategic and tactical objectives. The key to creating viable models is to populate the simulations with data that is truly representative of a group of people to the greatest extent possible. This thesis looked at survey data and narratives as methods of representing a society. Furthermore, this thesis examined the potential of using particular tools to process and prepare these types of data in order to minimize the resources and cost accrued through analyst input often required in the preparation and set-up of the input tables used to instantiate the CG model. These tools were the Survey Data Development Tool (or the Survey Data Case File Generator) developed by TRAC-MTRY and Sandia's Text ANaLysis Extensible librarY (STANLEY) document comparison tool.

The overarching question that this thesis attempted to answer involved the effectiveness of automating the process of agent population in social simulations. Both tools were analyzed with this question in mind, and to varying extents, both tools showed promise in this area. The survey data tool successfully produced case files that could instantiate the CG model. STANLEY, while producing results that were less encouraging due to sample size limitations, showed potential as a method of sentiment

forecasting and, through some additional experimentation and focused study, could be an effective implement to aid analysts in instantiating agents' beliefs and issue stances in social simulations.

#### **A. SURVEY DATA DEVELOPMENT TOOL CONCLUSIONS AND RECOMMENDATIONS**

The survey tool, described in II.B, develops case files from survey data for use in the CG model. The tool reads surveys, ranks survey questions or responses based on a target question in order of significance, and writes case files based on user definitions of particular terms. The tool was a work in progress from the beginning, but after early testing, and the aforementioned changes made by Mr. Greg Sullivan, the tool proved capable of developing case files with relative ease. Obviously there may be some modification of the survey data itself to meet the requirements of the tool or the Bayesian network generation software (Netica or Weka), but the tool was able to provide suitable case files that were successful in instantiating the CG model. The model behaved as expected, validating the case files and thus the tool.

#### **B. FUTURE WORK WITH THE SURVEY DATA DEVELOPMENT TOOL**

While this research focused primarily on the ability of the tool to produce case files for use in social simulations, and in particular the CG model, in addition to those mentioned throughout the thesis, there are several other potential areas of research and testing that can be done on and with the tool to further its uses.

- The tool ranks the terms based on Chi-square and variance, depending on data type. Are these the best methods for comparing results? Is there another analytical comparison that can be used? The responses in the surveys used in this thesis were string responses. How does the tool behave with respect to scoring and ranking if the survey data is mixed numerical and string? This would require some recoding of the source material for the tool, but if a more efficient and effective method of scoring the survey results can be found, this could perhaps increase the effectiveness of the social simulation output.
- It is critical that the tool be used in an application of the CG model using the full extent of CG's capabilities. A discussion of the full abilities of CG was beyond the scope of this study, but needless to say it is a powerful and competent tool capable of handling a variety of inputs to affect agent

behaviors. The assumptions made to hold model behavior distributions constant in order to gain deterministic results and that only changes to the three indicators (oil prices, unemployment, etc.) would affect agent behavior were made to limit model in its reactions and output and only check the affect of the tool generated case files. While this thesis' main focus was to determine if the tool could produce suitable case files, and a level of verification was done to that effect, a fuller and more robust test of the CG model will further validate the tool as a viable alternative to SME utilization in case file development. Furthermore, a cost benefit analysis should be accomplished to quantify the benefits automating the case file generation process.

- Since there was some disparity at the initial period of the model runs with the actual Pew results attributed to a possible warm up period, it is advisable to rerun the scenario with a 30–60 day acclimation period for the agents. Once that period is over, start the simulation at day Zero and check the output against what was analyzed in this research.
- At the time of this writing, the tool was still undergoing modification. Mr. John Ruck, a software developer for Rolands and Associates contracted to TRAC-MTRY, was working on a change to the code and interface to allow the user to select the order of the stereotypes in the naming of the case files. This is important because order matters in case file names in the CG model. All future uses of the tool should include these changes.

### **C. STANLEY TOOL CONCLUSIONS AND RECOMMENDATIONS**

Previous work with STANLEY (McKaughan, Alt, Heath, & McClain, 2011) yielded results that were less than desirable, but demonstrated the potential of the tool given that precise and specific classifiers were developed and utilized. This thesis took a different approach with the tool, and one that would appear to be more in line with its potential use in the population of social simulations than the previous study. In fact, TRAC-MTRY, in conjunction with Sandia National Laboratories, is investigating the total automation of STANLEY to provide sentiment analysis without the use of seed documents and classifiers. Until that is explored and realized, the research results indicate that it can be efficient to seed a web spider with a series of documents and a list of websites (the classifiers) and rely on the analysis tool that rides underneath the spider to handle the scoring and categorization of the results. This semi-automation of the process, as analyzed here, seems promising. The tool was able to show a correlation between the events of the “Arab Spring” and the increase in domestic oil prices. The

correlation values and regression results were not strong enough, nor were they backed up by sufficient sample sizes, for any definitive statement to be made regarding a relationship, but if certain thresholds are established to limit the range of documents returned and the limitations of the tool noted herein are corrected, then specific and appropriate documents corpus can be generated that will beneficially instantiate Bayesian networks and populate social simulations.

#### **D. FUTURE WORK WITH STANLEY**

This thesis examined the text analysis tool in a limited and highly structured manner. The scope of the classifiers was quite narrow, and the limitations of the tool with respect to date extraction were quite significant (and not known at the outset of the thesis research). Yet the tool showed promise. Therefore, there are several possible areas of future work with this tool.

- Date extraction may or may not be important for any future work with this tool. Regardless, since there is so much information available through the use of a web spider, it makes sense to be able to accurately date that data for whatever purposes the investigators might need. In the development of issue stances and population sentiment, it would certainly be beneficial to see how those stances change over time. Currently, SME input is required to develop those BVIs. Automating that process would save valuable time and resources.
- An investigation into the actual causes of similarity scores of zero should be accomplished to validate the assumptions made in this thesis. The difference in the correlation and R-squared values between the different sets of STANLEY results varied by as much as nearly 300% in some cases. This is significant and can greatly influence the effect of the tool on Bayesian network creation and agent instantiation.
- Looking at a larger application, in the same vein as this research, might prove the concept initially sought after here. Perhaps a study that examined the period of President Obama's term in office, and how various elections played out as a result. For example, the mid-term elections of 2010 saw a significant shift in power from the Democrats to the Republicans, especially in the House of Representatives. A comparable study to this thesis could be conducted to see if STANLEY could predict the results of that election. A 20 month archive of narratives gained from the Internet and elsewhere from January 2009 to November 2010 would ideally yield more robust results.

- Obviously the reason that TRAC-MTRY and the U.S. Army are looking at STANLEY is to ultimately be able to use it in issue stance prediction to create Bayesian networks to populate the CG model. The research definitely needs to extend to that area. A region specific scenario should be tested where narratives of particular populations of interest are analyzed and scored, the results fed into a Bayesian network generating tool such as Weka or Netica, and the created networks used to instantiate agents in a social simulation such as the CG model. Using STANLEY to create Bayesian networks, while briefly touched upon by McKaughan, et al (2011), has not been explored deeply enough and needs to be for the tool to prove its worth.
- After the fact, it was obvious that to get a firmer handle on the intricacies of how STANLEY operated, and to have a more direct say in how the research and experimentation were conducted, a visit to Sandia's lab should have been made. It is highly recommended that any future research team travel to Sandia's laboratory in Albuquerque, NM to get hands-on with the tool, to learn how it is set up and run; or coordinate a visit from Sandia's personnel to visit and conduct a live demonstration of the tool. To truly be able to harness the power and benefits of the tool, any potential user must be better educated on its range of uses, its capabilities, and its shortcomings.

THIS PAGE INTENTIONALLY LEFT BLANK

## APPENDIX A. SAMPLE PEW RESEARCH CENTER SURVEY

PEW RESEARCH CENTER FOR THE PEOPLE & THE PRESS

MARCH 2010 POLITICAL SURVEY

FINAL QUESTIONNAIRE

March 10, 2010

### LANDLINE INTRO:

Hello, I am \_\_\_\_\_ calling for SRBI Survey Research. We are conducting a telephone opinion survey for leading newspapers and TV stations around the country. I'd like to ask a few questions of the [RANDOMIZE: ~~YOUNGEST~~ MALE, 18 years of age or older, who is now at home" AND ~~YOUNGEST~~ FEMALE, 18 years of age or older, who is now at home?"] [IF NO MALE/FEMALE, ASK: May I please speak with the YOUNGEST FEMALE/MALE, 18 years of age or older, who is now at home?] GO TO MAIN INTERVIEW

### CELL PHONE INTRO:

Hello, I am \_\_\_\_ calling for SRBI Survey Research. We are conducting a survey for leading newspapers and TV stations around the country. I know I am calling you on a cell phone, as a small token of our appreciation for your time, we will pay all eligible respondents \$5 for participating in this survey.

This is not a sales call. (IF R SAYS DRIVING/UNABLE TO TAKE CALL; Thank you. We will try you another time...).

VOICE MAIL MESSAGE (LEAVE ONLY ONCE -- THE FIRST TIME A CALL GOES TO VOICEMAIL): I am calling for SRBI Survey Research. We are conducting a short national survey of cell phone users. This is NOT a sales call. We will try to reach you again.

### SCREENING INTERVIEW:

S1. Are you under 18 years old, OR are you 18 or older?

1 Under 18

2 18 or older

9 Don't know/Refused

IF S1=2, CONTINUE WITH MAIN INTERVIEW

IF S1=1,9 THANK AND TERMINATE: This survey is limited to adults age 18 and over. I won't take any more of your time...

READ TO ALL CELL PHONE

INTRODUCTION TO MAIN INTERVIEW: We're interested in learning more about people with cell phones. This will only take about 20 minutes. If you are now driving a car or doing any activity requiring your full attention, I need to call you back later. The first question is...

INTERVIEWER:

IF R SAYS IT IS NOT A GOOD TIME, TRY TO ARRANGE A TIME TO CALL BACK. OFFER THE TOLL-FREE CALL-IN NUMBER THEY CAN USE TO COMPLETE THE SURVEY BEFORE ENDING THE CONVERSATION.

RANDOMIZE Q.1 AND Q.2

ASK ALL:

Q.1 Do you approve or disapprove of the way Barack Obama is handling his job as President? [IF DK ENTER AS DK. IF DEPENDS PROBE ONCE WITH: Overall do you approve or disapprove of the way Barack Obama is handling his job as President? IF STILL DEPENDS ENTER AS DK] {2-10} {QID:PRESAPP}

- 1 Approve
- 2 Disapprove
- 9 Don't know/Refused (VOL.)

RANDOMIZE Q.1 AND Q.2

ASK ALL:

Q.2 All in all, are you satisfied or dissatisfied with the way things are going in this country today? {2-10} {QID:x931024-3}

- 1 Satisfied
- 2 Dissatisfied
- 9 Don't know/Refused (VOL.)

ASK FORM 1 ONLY:

Q.3F1 What one word best describes your impression of Congress these days? Just the first word that comes to mind? [OPEN END; PROBE ONCE IF RESPONDENT ANSWERS ~~“DON'T KNOW.”~~ ACCEPT MULTIPLE WORD RESPONSES BUT DO NOT PROBE FOR SECOND RESPONSE] {new} {QID:qid20100301q3f1}

- 1 Answer given
- 9 Don't know/Refused

RANDOMIZE Q.4 AND Q.5

ASK ALL:

Q.4 Do you approve or disapprove of the job the Republican leaders in Congress are doing?  
 [IF DK ENTER AS DK. IF DEPENDS PROBE ONCE WITH: Overall do you approve or disapprove of  
 the job the Republican leaders in Congress are doing? IF STILL DEPENDS ENTER AS DK] {1–10}  
 {QID:REPAPP}

- 1 Approve
- 2 Disapprove
- 9 Don't know/Refused (VOL.)

RANDOMIZE Q.4 AND Q.5

ASK ALL:

Q.5 Do you approve or disapprove of the job the Democratic leaders in Congress are doing?  
 [IF DK ENTER AS DK. IF DEPENDS PROBE ONCE WITH: Overall do you approve or disapprove of  
 the job the Democratic leaders in Congress are doing? IF STILL DEPENDS ENTER AS DK] {1–10}  
 {QID:DEMAPP}

- 1 Approve
- 2 Disapprove
- 9 Don't know/Refused (VOL.)

NO QUESTIONS 6–8

ASK ALL:

Thinking about some issues...

Q.9 Do you approve or disapprove of the way Barack Obama is handling [INSERT ITEM,  
 RANDOMIZE; OBSERVE FORM SPLITS] How about [NEXT ITEM]? [REPEAT INTRODUCTION AS  
 NECESSARY]

- a. The economy {1–10} {QID:PRESECON}
- b. Health care policy {1–10} {QID:PRESHEALTH}
- c.F1 The nation's foreign policy {1–10} {QID:PRESFP}
- d.F2 The situation in Afghanistan {1–10} {QID:PRESAFGHANISTAN}

RESPONSE CATEGORIES:

- 1 Approve
- 2 Disapprove
- 9 Don't know/Refused (VOL.)

ASK ALL:

Q.10 As I name some traits, please tell me whether you think each one describes Barack Obama. First, [INSERT FIRST ITEM; RANDOMIZE] do you think of Barack Obama as [FIRST ITEM; RANDOMIZE] or not? Do you think of Barack Obama as [NEXT ITEM] or not? {trend for comparison to 11–08 voters and earlier 08 trends on RVs}

- a. Inspiring {11–08 voters, 10–08, 9–08, 4–08, 3–08 all RVs} {QID:qid45190}
- b. Arrogant {9–08, 4–08 RVs} {QID:qid20080401q15g}
- c. Decisive {new} {QID:qid20100301q10c}
- d. Detached {new} {QID:qid20100301q10d}

RESPONSE CATEGORIES:

- 1 Yes, describes Barack Obama
- 2 No
- 9 Don't Know/Refused (VOL.)

ASK ALL:

Q.11 Does Barack Obama make you feel [INSERT FIRST ITEM; RANDOMIZE] or not? Does Obama make you feel [NEXT ITEM] or not? {trend for comparison to 11–08 voters, mod 3–08 RVs}

- a. Hopeful {11–08 voters, 3–08 RVs} {QID:qid20081102q67a}
- b. Proud {11–08 voters, 3–08 RVs} {QID:qid20081102q67b}
- c. Angry {11–08 voters, 3–08 RVs} {QID:qid20081102q67c}
- d. Disappointed {new} {QID:qid20100301q11d}

RESPONSE CATEGORIES:

- 1 Yes, makes you feel this way
- 2 No
- 9 Don't Know/Refused (VOL.)

NO QUESTIONS 12–13

ASK ALL:

Q.14 In dealing with important issues facing the country, are [RANDOMIZE; Democrats in Congress and Republicans in Congress] working together or not working together? {new} {QID:qid20100301q14}

- 1 Working together
- 2 Not working together
- 9 Don't know/Refused (VOL.)

ASK IF 'NOT WORKING TOGETHER' (Q.14=2):

Q.14a Who do you think is most to blame for them not working together? [READ AND RANDOMIZE IN SAME ORDER AS Q.14] {new} {QID:qid20100301q14a}

- 1 Democratic leaders in Congress [OR]
- 2 Republican leaders in Congress
- 3 [VOL. DO NOT READ] Both
- 4 [VOL. DO NOT READ] Neither
- 5 [VOL. DO NOT READ] Other (SPECIFY: \_\_\_\_\_)
- 9 [VOL. DO NOT READ] Don't know/Refused

ASK ALL:

Q.15 Which political party has the higher moral and ethical standards? [READ AND RANDOMIZE] {new} {QID:qid20100301q15}

- 1 The Democratic Party [OR]
- 2 The Republican Party
- 3 [VOL. DO NOT READ] Both
- 4 [VOL. DO NOT READ] Neither
- 9 [VOL. DO NOT READ] Don't know/Refused

ASK ALL:

On another subject ...

Q.16 What do you think is more important – to protect the right of Americans to own guns, OR to control gun ownership? {4-09} {QID:x990516-19}

- 1 Protect the right of Americans to own guns
- 2 Control gun ownership
- 9 Don't know/Refused (VOL.)

NO QUESTIONS 17-18

RANDOMIZE Q.19-Q.20/Q.22-Q.23 IN BLOCKS:

ASK ALL:

Thinking now about the nation's economy...

Q.19 How would you rate economic conditions in this country today... as excellent, good, only fair, or poor? {2-10} {QID:x040425-C1}

- 1 Excellent
- 2 Good
- 3 Only fair
- 4 Poor
- 9 Don't know/Refused (VOL.)

ASK ALL:

Q.20 Which of the these best describes your opinion: [READ; READ CATEGORIES IN REVERSE ORDER FOR HALF THE SAMPLE]? {new} {QID:qid20100301q20}

- 1 The economy is now recovering [OR]
- 2 The economy is not yet recovering but will recover soon [OR]
- 3 It will be a long time before the economy recovers
- 9 [VOL. DO NOT READ] Don't know/Refused

NO QUESTION 21

RANDOMIZE Q.19-Q.20/Q.22-Q.23 IN BLOCKS:

ASK ALL:

Thinking now about your own personal finances...

Q.22 How would you rate your own personal financial situation? Would you say you are in excellent shape, good shape, only fair shape or poor shape financially? {12-09} {QID:x931205-14}

- 1 Excellent shape
- 2 Good shape
- 3 Only fair shape
- 4 Poor shape
- 9 Don't know/Refused (VOL.)

ASK ALL:

Q.23 Over the course of the next year, do you think the financial situation of you and your family will improve a lot, improve some, get a little worse or get a lot worse? {12-09} {QID:x940727-26}

- 1 Improve a lot
- 2 Improve some
- 3 Get a little worse
- 4 Get a lot worse
- 5 Stay the same (VOL.)
- 9 Don't know/Refused (VOL.)

ASK ALL:

Q.24 Which of the following national economic issues worries you most? [READ AND RANDOMIZE] {new} {QID:qid20100301q24}

- 1 Problems in the financial and housing markets
- 2 The federal budget deficit
- 3 Rising prices [OR]
- 4 The job situation
- 7 [VOL. DO NOT READ] Other
- 8 [VOL. DO NOT READ] None / not worried about any
- 9 [VOL. DO NOT READ] Don't know/Refused

ASK ALL:

Now thinking about health care...

Q.25 How much, if anything, have you heard about the bills in Congress to overhaul the health care system? Have you heard...[READ] {2-10} {QID:qid20090701q48}

- 1 A lot
- 2 A little [OR]
- 3 Nothing at all
- 9 [VOL. DO NOT READ] Don't know/Refused

ASK ALL:

Q.26 As of right now, do you generally favor or generally oppose the health care bills being discussed in Congress? {2-10} {QID:qid20090701q49}

- 1 Generally favor
- 2 Generally oppose
- 9 Don't know/Refused (VOL.)

ASK THOSE WHO OPPOSE HEALTH CARE BILLS (Q.26=2):

Q.27 Would you prefer that Congress [INSERT OPTION, RANDOMIZE] or would you prefer that Congress [NEXT OPTION]? {new, trend for comparison to 2-10?} {QID:qid20100301q27}

- 1 Begin working on new health care legislation
- 2 Pass nothing and leave the current system as it is
- 9 [VOL. DO NOT READ] Don't know/Refused

NO QUESTIONS 28-30

ASK FORM 1 ONLY:

Q.31F1 If the health care bills being discussed in Congress are passed into law, do you think your health care costs would [RANDOMIZE: increase, decrease] or stay the same in the coming years? [IF INCREASE: Do you think they would increase a lot or a little?] {new} {QID:qid20100301q31f1}

- 1 Increase, a lot
- 2 Increase, a little
- 3 Increase, don't know how much (VOL.)
- 4 Decrease
- 5 Stay the same
- 9 Don't know/Refused (VOL.)

ASK FORM 2 ONLY:

Q.32F2 If no changes are made to the health care system, do you think your health care costs would [RANDOMIZE: increase, decrease] or stay the same in the coming years? [IF INCREASE: Do you think they would increase a lot or a little?] {new} {QID:qid20100301q32f2}

- 1 Increase, a lot
- 2 Increase, a little
- 3 Increase, don't know how much (VOL.)
- 4 Decrease
- 5 Stay the same

9 Don't know/Refused (VOL.)

ASK ALL:

HEALTHINS Are you, yourself, now covered by any form of health insurance or health plan or are you not covered at this time? [READ IF NECESSARY: A health plan would include any private insurance plan through your employer or a plan that you purchased yourself, as well as a government program like Medicare or Medicaid] {1-10} {QID:qid20090601q59}

- 1 Covered by health insurance
- 2 Not covered by health insurance
- 9 Don't know/Refused (VOL.)

ASK ALL:

EMPLOY Are you now employed full-time, part-time or not employed? {12-09}  
{QID:EMPLOY}

- 1 Full-time
- 2 Part-time
- 3 Not employed
- 9 Don't know/Refused (VOL.)

ASK IF NOT EMPLOYED (EMPLOY=3):

EMPLOY1 Are you currently looking for work, or not? {2-09} {QID:EMPLOY1}

- 1 Yes, looking for work
- 2 No, not looking
- 9 Don't know/Refused (VOL.)

ASK ALL:

STUDENT Are you now enrolled in school, either full or part-time, or not? {5-09}  
{QID:STUDENT}

- 1 Yes, full-time student
- 2 Yes, part-time
- 3 No
- 9 Don't know/Refused (VOL.)

NO QUESTIONS 33-35

ASK ALL:

Q.36 Thinking now about job opportunities where you live, would you say there are plenty of jobs available in your community or are jobs difficult to find? {10-09} {QID:x010617-20}

- 1 Plenty of jobs available
- 2 Jobs are difficult to find
- 3 Lots of some jobs, few of others (VOL.)
- 9 Don't know/Refused (VOL.)

ASK IF EMPLOYED (EMPLOY=1,2) OR NOT EMPLOYED AND NOT CURRENTLY LOOKING FOR WORK (EMPLOY1=2,9)

Q.37 Over the past 12 months, has there been a time when you or someone in your household has been without a job and looking for work, or not? {10-09} {QID:x041216-50d}

- 1 Yes
- 2 No
- 9 Don't know/Refused (VOL.)

ASK ALL:

Q.38 For each of the following, please tell me whether or not it is something that happened to you in the past year....Have you [INSERT ITEM; RANDOMIZE ITEMS a. THRU f. WITH ITEM g. ALWAYS LAST]? [IF RESPONDENT SAYS THIS DOES NOT APPLY, CODE AS NO] {2-09, 1-08 SDT, 8-99 Post/Kaiser}

a. Had trouble getting or paying for medical care for yourself or your family {2-09, 1-08, 8-99} {QID:qid20090201q39a}

b. Had problems paying your rent or mortgage {2-09, 1-08, 8-99} {QID:qid20090201q39b}

c. Been laid off or lost your job {2-09, 1-08} {QID:qid20090201q39c}

d. Gotten a pay raise at your current job or gotten a better job {2-09, 1-08, 8-99} {QID:qid20090201q39d}

e. Had a mortgage, other loan, or credit card application denied {2-09} {QID:qid20090201q39e}

f. Had problems with collection or credit agencies {2-09, 8-99} {QID:qid20090201q39f}

ASK IF NOT LAID OFF (Q.38c=2,9):

g. Had your hours reduced or been asked to take a cut in pay {2-09} {QID:qid20090201q39g}

RESPONSE CATEGORIES:

- 1 Yes
- 2 No
- 9 Don't know/Refused (VOL.)

ASK IF EMPLOYED (EMPLOY=1,2):

Q.39 Thinking about the next 12 months, how likely is it that [INSERT FIRST ITEM; RANDOMIZE]? Is it very likely, somewhat likely, not too likely or not at all likely? What about [INSERT NEXT ITEM]? [IF NECESSARY: Is it very likely, somewhat likely, not too likely or not at all likely that (REPEAT ITEM)] {2-09}

- a. Your employer may go out of business or relocate to another city {2-09, 1-08} {QID:qid20090201q40a}
- b. You may have your health care benefits reduced or eliminated by your employer {2-09, 1-08} {QID:qid20090201q40c}
- c. You may be asked to take a cut in pay {2-09, 1-08} {QID:qid20090201q40d}
- d. You may be laid off {2-09, 1-08} {QID:qid20090201q40f}
- e. You may have your retirement benefits reduced or eliminated by your employer {2-09} {QID:qid20090201q40g}

RESPONSE CATEGORIES:

- 1 Very likely
- 2 Somewhat likely
- 3 Not too likely
- 4 Not at all likely
- 9 Don't know/Refused (VOL.)

NO QUESTIONS 40-43

ASK ALL:

Q.44 Thinking about gun laws... do you think state and local governments should or should not be able to pass laws that ban the possession or sale of handguns in their jurisdictions? {new} {QID:qid20100301q44}

- 1 Should
- 2 Should not
- 9 Don't know/Refused (VOL.)

ASK ALL:

Thinking about Afghanistan...

Q.45 How well is the U.S. military effort in Afghanistan going? [READ IN ORDER] {12-09}  
{QID:qid20080202Q80}

- 1 Very well
- 2 Fairly well
- 3 Not too well
- 4 Not at all well
- 9 [VOL. DO NOT READ] Don't know/Refused

ASK ALL:

Q.46 Regardless of what you think about the original decision to use military force in Afghanistan, do you now believe that the United States will definitely succeed, probably succeed, probably fail, or definitely fail in achieving its goals in Afghanistan? {12-09} {QID:qid20090101q53}

- 1 Definitely succeed
- 2 Probably succeed
- 3 Probably fail
- 4 Definitely fail
- 9 Don't know/Refused (VOL.)

NO QUESTIONS 47-48

ASK ALL:

Now thinking about Iraq...

Q.49 How well is the U.S. military effort in Iraq going? [READ IN ORDER] {12-09}  
{QID:x030327-PRC5}

- 1 Very well
- 2 Fairly well
- 3 Not too well
- 4 Not at all well
- 9 [VOL. DO NOT READ] Don't know/Refused

ASK ALL:

Q.50 Regardless of what you think about the original decision to use military force in Iraq, do you now believe that the United States will definitely succeed, probably succeed, probably fail, or definitely fail in achieving its goals in Iraq? {12-09} {QID:x060910-63F2}

- 1 Definitely succeed
- 2 Probably succeed
- 3 Probably fail
- 4 Definitely fail
- 9 Don't know/Refused (VOL.)

ASK ALL:

On a different subject ...

QCEN.1 Have you ever heard of the United States Census, or have you not heard of this? {1-10}  
{QID:qid20100101QC.1}

- 1 Yes, have heard
- 2 No, have not heard
- 9 Don't know/Refused (VOL.)

ASK IF HAVE NOT HEARD OF THE CENSUS (QCEN.1=2,9):

QCEN.1a The Census is the count of all people who live in the United States. Have you ever heard of that before, or have you not heard of that? {1-10} {QID:qid20100101QC.2}

- 1 Yes, have heard
- 2 No, have not heard
- 9 Don't know/Refused (VOL.)

ASK IF HEARD OF THE CENSUS (QCEN.1=1 OR QCEN1a=1):

On a different subject ...

QCEN.2 Have you seen or heard anything recently—within the last month or so—about the Census, or have you not seen or heard anything recently? {1-10} {QID:qid20100101QC.2}

- 1 Yes, have seen or heard recently
- 2 No, have not seen or heard recently
- 9 Don't know/Refused (VOL.)

ASK ALL:

QCEN.3 How likely are you to participate in the 2010 Census? By participate, we mean fill out and mail in a Census form. Would you say you [READ]? [IF RESPONDENT SAYS THEY ALREADY RECEIVED AND FILLED OUT OR MAILED IN THE FORM, PLEASE CODE AS PUNCH 6] {1-10, slight mod to options 6 and 7} {QID:qid20100101QC.4}

- 1 Definitely will
- 2 Probably will
- 3 Might or might not
- 4 Probably will not
- 5 Definitely will not
- 6 [VOL. DO NOT READ] Already filled out and mailed in form
- 7 [VOL. DO NOT READ] Other, (SPECIFY: \_\_\_\_\_)
- 9 [VOL. DO NOT READ] Don't know/Refused

ASK IF MIGHT OR MIGHT NOT OR WILL NOT PARTICIPATE (QCEN.3=3,4,5):

QCEN.4 How likely is it that someone else in your household will participate in the 2010 Census? Would you say someone else in your household [READ]? [IF NECESSARY: By participate, we mean fill out and mail in a Census form.] [IF RESPONDENT SAYS SOMEBODY IN THEIR HOUSEHOLD ALREADY RECEIVED AND FILLED OUT OR MAILED IN THE FORM, PLEASE CODE AS PUNCH 6] {new} {QID:qid20100301qcen4}

- 1 Definitely will
- 2 Probably will
- 3 Might or might not
- 4 Probably will not
- 5 Definitely will not
- 6 [VOL. DO NOT READ] Already filled out and mailed in form
- 7 [VOL. DO NOT READ] Other, (SPECIFY: \_\_\_\_\_)
- 9 [VOL. DO NOT READ] Don't know/Refused

ASK ALL:

QCEN.5 How important do you think the Census is for the United States? Would you say it is [READ]? {1-10} {QID:qid20100101QC.5}

- 1 Very important
- 2 Somewhat important
- 3 Not too important
- 4 Not at all important
- 9 [VOL. DO NOT READ] Don't know/Refused

ASK ALL:

QCEN.6 Do you believe that answering and sending back your Census form would personally benefit you in any way, personally harm you, or neither benefit nor harm you? {1–10, slight mod to add option 4} {QID:qid20100101QC.6}

- 1 Personally benefit
- 2 Personally harm
- 3 Neither benefit nor harm
- 4 Both benefit and harm (VOL.)
- 9 Don't know/Refused (VOL.)

ASK ALL:

QCEN.7 Do you believe that filling out your Census form would benefit your community, harm your community, or neither benefit nor harm your community? {new} {QID:qid20100301qcen7}

- 1 Benefit
- 2 Harm
- 3 Neither benefit nor harm
- 4 Both benefit and harm (VOL.)
- 9 Don't know/Refused (VOL.)

RANDOMIZE ORDER OF QCEN.8 THROUGH QCEN.15:

ASK ALL:

QCEN.8 Do you think filling out the Census form will take too much time, or don't you think so? {new} {QID:qid20100301qcen8}

- 1 Yes, will take too much time
- 2 No, don't think so
- 9 Don't know/Refused (VOL.)

RANDOMIZE ORDER OF QCEN.8 THROUGH QCEN.15:

ASK ALL:

QCEN.9 Do you think conducting the 2010 Census will cost too much money, or don't you think so? {new} {QID:qid20100301qcen9}

- 1 Yes, will cost too much money
- 2 No, don't think so
- 9 Don't know/Refused (VOL.)

RANDOMIZE ORDER OF QCEN.8 THROUGH QCEN.15:

ASK ALL:

QCEN.10 Do you think the Census Bureau will keep personal information confidential, or don't you think so? {new} {QID:qid20100301qcen10}

- 1 Yes, will keep personal information confidential
- 2 No, don't think so
- 9 Don't know/Refused (VOL.)

RANDOMIZE ORDER OF QCEN.8 THROUGH QCEN.15:

ASK ALL:

QCEN.11 Do you think the Census asks for more personal information than the government really needs, or don't you think so? {new} {QID:qid20100301qcen11}

- 1 Yes, asks for more personal information than the government really needs
- 2 No, don't think so
- 9 Don't know/Refused (VOL.)

RANDOMIZE ORDER OF QCEN.8 THROUGH QCEN.15:

ASK ALL:

QCEN.12 Do you think participating in the Census is a civic responsibility, or don't you think so? {new} {QID:qid20100301qcen12}

- 1 Yes, participating in the Census is a civic responsibility
- 2 No, don't think so
- 9 Don't know/Refused (VOL.)

RANDOMIZE ORDER OF QCEN.8 THROUGH QCEN.15:

ASK ALL:

QCEN.13 Do you think the results of the Census help one political party more than the other, or don't you think so? {new} {QID:qid20100301qcen13}

- 1 Yes, helps one political party more than the other
- 2 No, don't think so
- 9 Don't know/Refused (VOL.)

ASK IF YES (QCEN.13=1):

QCEN.13a Which political party do you think the Census helps more? [OPEN END. SINGLE RESPONSE. USE PRECODES AS APPROPRIATE. ACCEPT ONLY ONE RESPONSE.] {new} {QID:qid20100301qcen13a}

(PRECODES – DO NOT READ)

- 1 Republican Party
- 2 Democratic Party
- 3 Other, (SPECIFY: \_\_\_\_\_)
- 4 None
- 9 Don't know/Refused

RANDOMIZE ORDER OF QCEN.8 THROUGH QCEN.15:

ASK ALL:

QCEN.14 As far as you know, is the Census used to decide how many representatives each state will have in Congress, or is it not used for this? {1–10} {QID:qid20100101QC.8}

- 1 Yes, used for this
- 2 No, not used for this
- 9 Don't know/Refused (VOL.)

RANDOMIZE ORDER OF QCEN.8 THROUGH QCEN.15:

ASK ALL:

QCEN.15 As far as you know, is the Census used to determine whether someone is in this country legally, or is it not used for this? {new, for comparison to Pew Hispanic March Omnibus} {QID:qid20100301qcen15}

- 1 Yes, used for this
- 2 No, not used for this
- 9 Don't know/Refused (VOL.)

ASK ALL:

Now thinking about another topic...

Q.52 Do you think the use of marijuana should be made legal, or not? {new, Gallup 10–09} {QID:qid20100301q52}

- 1 Yes, legal

- 2 No, illegal
- 9 Don't know/Refused (VOL.)

ASK ALL:

Q.53 Do you favor or oppose your state allowing the sale and use of marijuana for medical purposes if it is prescribed by a doctor? {new} {QID:qid20100301q53}

- 1 Favor
- 2 Oppose
- 9 Don't know/Refused (VOL.)

ASK ALL:

Q.54 Do you think that allowing medical marijuana makes it easier for people to get marijuana even if they don't have a real medical need, or doesn't it make a difference? {new} {QID:qid20100301q54}

- 1 Yes, makes it easier
- 2 Doesn't make a difference
- 9 Don't know/Refused

ASK IF EASIER (Q.54=1):

Q.55 Are you very concerned, somewhat concerned, not too concerned, or not at all concerned about this? [READ] [REPEAT IF NECESSARY: That allowing medical marijuana makes it easier for people to get marijuana even if they don't have a real medical need?] {new} {QID:qid20100301q55}

- 1 Very concerned
- 2 Somewhat concerned
- 3 Not too concerned
- 4 Not at all concerned
- 9 [VOL. DO NOT READ] Don't know/Refused

ASK ALL:

Q.56 How concerned, if at all, would you be if a store that sold medical marijuana opened up near other stores in your area? [READ] {new} {QID:qid20100301q56}

- 1 Very concerned
- 2 Somewhat concerned

- 3 Not too concerned
- 4 Not at all concerned
- 9 [VOL. DO NOT READ] Don't know/Refused

ASK ALL:

Q.57 Keeping in mind that all of your answers in this survey are confidential, have you, yourself, ever happened to try marijuana? {6-03} {QID:x010219-41}

- 1 Yes
- 2 No
- 9 Don't know/Refused (VOL.)

ASK ALL:

Now, just a few questions for statistical purposes only. {QID:SEX}

SEX [ENTER RESPONDENT'S SEX:]

- 1 Male
- 2 Female

ASK ALL:

AGE What is your age? {QID:AGE}

\_\_\_\_\_ years

- 97 97 or older
- 99 Don't know/Refused (VOL.)

ASK ALL:

EDUC What is the last grade or class that you completed in school? [DO NOT READ] {QID:EDUC}

- 1 None, or grade 1-8
- 2 High school incomplete (Grades 9-11)
- 3 High school graduate (Grade 12 or GED certificate)
- 4 Technical, trade, or vocational school AFTER high school
- 5 Some college, associate degree, no 4-year degree
- 6 College graduate (B.S., B.A., or other 4-year degree)
- 7 Post-graduate training or professional schooling after college

(e.g., toward a master's Degree or Ph.D.; law or medical school)

9 Don't know/Refused (VOL.)

ASK RANDOM ONE HALF-SAMPLE HISP2, OTHER HALF-SAMPLE HISP3

[NOTE: DO NOT SPLIT BY FORM]:

HISP2 Are you Spanish, Hispanic, or Latino [PAUSE] or not? {QID:HISP2}

1 Yes

2 No

9 Don't know/Refused (VOL.)

ASK RANDOM ONE HALF-SAMPLE HISP2, OTHER HALF-SAMPLE HISP3

[NOTE: DO NOT SPLIT BY FORM]:

HISP3 Are you of Hispanic, Latino, or Spanish origin [PAUSE] or not? {QID:HISP3}

1 Yes

2 No

9 Don't know/Refused (VOL.)

ASK ALL:

RACE1 Which of the following describes your race? You can select as many as apply.  
{QID:RACE1}

[READ LIST. RECORD UP TO FOUR RESPONSES IN ORDER MENTIONED]

1 White

2 Black or African-American

3 Asian or Asian-American

4 Or some other race

9 [VOL. DO NOT READ] Don't know/Refused

ASK ALL:

MARITAL Are you currently married, living with a partner, divorced, separated, widowed,  
or have you never been married? [IF R SAYS ~~SINGLE~~," PROBE TO DETERMINE WHICH  
CATEGORY IS APPROPRIATE] {QID:MARITAL1}

1 Married

- 2 Living with a partner
- 3 Divorced
- 4 Separated
- 5 Widowed
- 6 Never been married
- 9 Don't know/Refused (VOL.)

ASK ALL:

HH1 How many people, including yourself, live in your household?

INTERVIEWER NOTE: HOUSEHOLD MEMBERS INCLUDE PEOPLE WHO THINK OF THIS HOUSEHOLD AS THEIR PRIMARY PLACE OF RESIDENCE, INCLUDING THOSE WHO ARE TEMPORARILY AWAY ON BUSINESS, VACATION, IN A HOSPITAL, OR AWAY AT SCHOOL. THIS INCLUDES INFANTS, CHILDREN AND ADULTS. {QID:HH1}

ENTER NUMBER (RANGE 1–50)

- 99 Don't know/Refused

ASK IF MORE THAN ONE PERSON IN HH (HH1=2–50):

HH2 How many of these are children under the age of 18? {QID:HH2}

ENTER NUMBER (RANGE 0–50)

- 99 Don't know/Refused

ASK ALL:

RELIG What is your present religion, if any? Are you Protestant, Roman Catholic, Mormon, Orthodox such as Greek or Russian Orthodox, Jewish, Muslim, Buddhist, Hindu, atheist, agnostic, something else, or nothing in particular? {QID:RELIG}

[INTERVIEWER: IF R VOLUNTEERS ~~nothing~~ nothing in particular, none, no religion, etc.” BEFORE REACHING END OF LIST, PROMPT WITH: And would you say that’s atheist, agnostic, or just nothing in particular?]

- 1 Protestant (Baptist, Methodist, Non-denominational, Lutheran, Presbyterian, Pentecostal, Episcopalian, Reformed, Church of Christ, Jehovah’s Witness, etc.)
- 2 Roman Catholic (Catholic)
- 3 Mormon (Church of Jesus Christ of Latter-day Saints/LDS)
- 4 Orthodox (Greek, Russian, or some other orthodox church)
- 5 Jewish (Judaism)

- 6 Muslim (Islam)
- 7 Buddhist
- 8 Hindu
- 9 Atheist (do not believe in God)
- 10 Agnostic (not sure if there is a God)
- 11 Something else (SPECIFY: \_\_\_\_\_)
- 12 Nothing in particular
- 13 Christian (VOL.)
- 14 Unitarian (Universalist) (VOL.)
- 99 Don't Know/Refused (VOL.)

ASK IF SOMETHING ELSE OR DK/REF (RELIG=11, 99):

CHR Do you think of yourself as a Christian or not?

IF R NAMED A NON-CHRISTIAN RELIGION IN PREVIOUS QUESTION (e.g., Native American, Wiccan, Pagan, etc.), DO NOT READ (ENTER "NO" CODE 2) {QID:CHR}

- 1 Yes
- 2 No
- 9 Don't know/Refused (VOL.)

ASK IF CHRISTIAN (RELIG=1-4, 13 OR CHR=1):

BORN Would you describe yourself as a "born again" or evangelical Christian, or not?  
{QID:BORN}

- 1 Yes, would
- 2 No, would not
- 9 Don't know/Refused (VOL.)

ASK ALL:

INCOME Last year, that is in 2009, what was your total family income from all sources, before taxes? Just stop me when I get to the right category. [READ] {QID:INCOME}

- 1 Less than \$10,000
- 2 10 to under \$20,000
- 3 20 to under \$30,000
- 4 30 to under \$40,000

- 5 40 to under \$50,000
- 6 50 to under \$75,000
- 7 75 to under \$100,000
- 8 100 to under \$150,000
- 9 \$150,000 or more
- 10 [VOL. DO NOT READ] Don't know/Refused

ASK ALL:

REGIST These days, many people are so busy they can't find time to register to vote, or move around so often they don't get a chance to re-register. Are you NOW registered to vote in your precinct or election district or haven't you been able to register so far? [INSTRUCTION: IF RESPONDENT VOLUNTEERS THAT THEY ARE IN NORTH DAKOTA AND DON'T HAVE TO REGISTER, PUNCH 1 FOR REGIST AND REGICERT] {QID:REGIST}

- 1 Yes, registered
- 2 No, not registered
- 9 Don't know/Refused (VOL.)

ASK IF RESPONDENT ANSWERED \_1\_ YES IN REGIST:

REGICERT Are you absolutely certain that you are registered to vote, or is there a chance that your registration has lapsed because you moved or for some other reason? {QID:REGICERT}

- 1 Absolutely certain
- 2 Chance registration has lapsed
- 9 Don't know/Refused (VOL.)

ASK ALL:

PARTY In politics TODAY, do you consider yourself a Republican, Democrat, or Independent? {QID:PARTY}

- 1 Republican
- 2 Democrat
- 3 Independent
- 4 No preference (VOL.)
- 5 Other party (VOL.)
- 9 Don't know/Refused (VOL.)

ASK IF ANSWERED 3, 4, 5 OR 9 IN PARTY:

PARTYLN As of today do you lean more to the Republican Party or more to the Democratic Party? {QID:PARTYLN}

- 1 Republican
- 2 Democrat
- 9 Other/Don't know/Refused (VOL.)

ASK ALL:

IDEO In general, would you describe your political views as... [READ] {QID:IDEO}

- 1 Very conservative
- 2 Conservative
- 3 Moderate
- 4 Liberal, OR
- 5 Very liberal?
- 9 [VOL. DO NOT READ] Don't know/Refused

ASK IF AGE>29:

OWNRENT Do you own or rent your home? {2-10 PST, 3-09 PP mod filter}  
{QID:OWNRENT}

- 1 Own
- 2 Rent
- 3 Other arrangement (VOL.)
- 9 Don't know/Refused (VOL.)

ASK IF AGE<30:

HOME2 Do you own your home, rent, live in a dorm or live with your parents? {2-10 PST}  
{QID:OWNRENT30}

- 1 Own
- 2 Rent
- 3 Live in a dorm
- 4 Live with parents
- 5 Other arrangement (VOL.)
- 9 Don't know/Refused (VOL.)

ASK IF HISPANIC (HISP2=1 or HISP3=1):

BIRTH\_HISP Were you born on the island of Puerto Rico, in the United States, or in another country?

{2-10; Do not show in Topline} {QID:BIRTH\_HISP}

- 1 Puerto Rico
- 2 U.S.
- 3 Another country
- 9 Don't know/Refused (VOL.)

ASK IF HISPANIC (HISP2=1 or HISP3=1) AND IF BORN IN U.S. (BIRTH\_HISP=2):

MOTHER\_HISP Was your mother born on the island of Puerto Rico, in the United States, or in another country? {2-10; Do not show in Topline}  
{QID:MOTHER\_HISP}

- 1 Puerto Rico
- 2 U.S.
- 3 Another country
- 9 Don't know/Refused (VOL.)

ASK IF HISPANIC (HISP2=1 or HISP3=1) AND IF BORN IN U.S. (BIRTH\_HISP=2):

FATHER\_HISP Was your father born on the island of Puerto Rico, in the United States, or in another country? {2-10;Do not show in Topline}  
{QID:FATHER\_HISP}

- 1 Puerto Rico
- 2 U.S.
- 3 Another country
- 9 Don't know/Refused (VOL.)

ASK ALL LANDLINE SAMPLE:

L1. Now thinking about your telephone use... Do you have a working cell phone? {QID:L1}

- 1 Yes, have cell phone

- 2 No, do not
- 9 Don't know/Refused (VOL.)

ASK IF NO CELL PHONE AND MULTI-PERSON HOUSEHOLD (L1=2,9 AND HH1>1):

L1a. Does anyone in your household have a working cell phone? {QID:L1a}

- 1 Yes, someone in household has cell phone
- 2 No
- 9 Don't know/Refused (VOL.)

ASK IF DUAL REACHED ON LANDLINE AND SINGLE-PERSON HOUSEHOLD (L1=1 AND HH1=1):

L2. Of all the telephone calls that you receive, do you get [READ AND RANDOMIZE OPTIONS 1 AND 3—KEEP 2 ALWAYS IN THE MIDDLE]? {QID:L2}

- 1 All or almost all calls on a cell phone
- 2 Some on a cell phone and some on a regular home phone
- 3 All or almost all calls on a regular home phone
- 9 [VOL. DO NOT READ] Don't know/Refused

ASK IF DUAL REACHED ON LANDLINE AND MULTI-PERSON HOUSEHOLD ((L1=1 OR L1a=1) AND HH1>1):

L3. Now thinking about all the people in your household, including yourself, of all the telephone calls that your household receives, are [READ AND RANDOMIZE OPTIONS 1 AND 3—KEEP 2 ALWAYS IN THE MIDDLE]? {QID:L3}

- 1 All or almost all calls on a cell phone
- 2 Some on a cell phone and some on a regular home phone
- 3 All or almost all calls on a regular home phone
- 9 [VOL. DO NOT READ] Don't know/Refused

ASK ALL CELL PHONE SAMPLE:

C1. Now thinking about your telephone use... Is there at least one telephone INSIDE your home that is currently working and is not a cell phone? {QID:C1}

- 1 Yes home telephone
- 2 No, home telephone

9 Don't know/Refused (VOL.)

ASK IF DUAL REACHED ON CELL PHONE AND SINGLE-PERSON HOUSEHOLD (C1=1 AND HH1=1):

C2. Of all the telephone calls that you receive, do you get [READ AND RANDOMIZE OPTIONS 1 AND 3—KEEP 2 ALWAYS IN THE MIDDLE]? {QID:C2}

- 1 All or almost all calls on a cell phone
- 2 Some on a cell phone and some on a regular home phone
- 3 All or almost all calls on a regular home phone
- 9 [VOL. DO NOT READ] Don't know/Refused

ASK IF DUAL REACHED ON CELL PHONE AND MULTI-PERSON HOUSEHOLD (C1=1 AND HH1>1):

C3. Now thinking about all the people in your household, including yourself, of all the telephone calls that your household receives, are [READ AND RANDOMIZE OPTIONS 1 AND 3—KEEP 2 ALWAYS IN THE MIDDLE]? {QID:C3}

- 1 All or almost all calls on a cell phone
- 2 Some on a cell phone and some on a regular home phone
- 3 All or almost all calls on a regular home phone
- 9 [VOL. DO NOT READ] Don't know/Refused

ASK ALL:

ZIPCODE What is your zipcode? {QID:ZIPCODE}

\_\_\_\_\_ Enter Zipcode

9 Don't know/Refused

END OF INTERVIEW

ASK ALL CELL PHONE SAMPLE:

MONEY That's the end of the interview. We'd like to send you \$5 for your time. Can I please have your full name and a mailing address where we can send you the money? [INTERVIEWER NOTE: If R does not want to give full name, explain we only need it so we can send the \$5 to them personally.] {QID:MONEY}

1 [ENTER FULL NAME] – INTERVIEWER: PLEASE VERIFY SPELLING

- 2 [ENTER MAILING ADDRESS]
- 3 [City]
- 4 [State]
- 5 CONFIRM ZIP from above
- 9 (VOL.) Respondent does not want the money

ASK ALL:

Thank you very much for your time. This survey is being conducted by the Pew Research Center for the People and the Press, which will be issuing a report on the results of this survey on their website, peoplepress dot ORG, in the coming weeks.

THANK YOU again for you help! Have a nice day/evening.

I HEREBY ATTEST THAT THIS IS A TRUE AND HONEST INTERVIEW.

INTERVIEWER GENDER:

ISEX {QID:ISEX}

- 1 Male
- 2 Female

INTERVIEWER RACE:

IHISP1 Are you, yourself, of Hispanic origin or descent, such as Mexican, Puerto Rican, Cuban, or some other Spanish background? {QID:IHISP1}

- 1 Yes
- 2 No
- 9 Don't know/Refused (VOL.)

IRACE1 Which of the following describes your race? You can select as many as apply.

[READ LIST. RECORD UP TO FOUR RESPONSES IN ORDER MENTIONED]  
{QID:IRACE1}

- 1 White
- 2 Black or African-American
- 3 Asian or Asian-American
- 4 Or some other race

[PLEASE MAKE THE FOLLOWING TEXT AVAILABLE TO INTERVIEWERS ANYTIME A RESPONDENT ASKS ABOUT THE NATURE OF THE PEW RESEARCH CENTER]The Pew Research Center for the People & the Press is an independent nonpartisan public opinion research organization that studies attitudes toward politics, the press and issues facing the nation. The Center has no connection to the government, political parties, or any campaigns. Reports about its surveys are made available free of charge on their website peoplepress dot ORG.

THIS PAGE INTENTIONALLY LEFT BLANK

## **APPENDIX B. PRELIMINARY WORK WITH STANLEY**

Prior to conducting the tests with oil prices, STANLEY was experimented with using a library of online works of literature. The purpose of this early analysis was to become familiar with the tool and develop a baseline standard procedure for set up and operation. The below text is an excerpt from the paper by McKaughan, Alt, Heath, and McClain (2011).

### **A. EARLY TESTS WITH STANLEY**

This section provides an illustrative application of the tool and a methodology to generate a Bayesian network describing the likelihood of a document in the corpus being classified as “accepting” or “unaccepting” of violence given its classification in three factors.

### **B. CASE STUDY**

This case study used Project Gutenberg to test the application of the STANLEY tool in the generation of a Bayesian Belief Network. Project Gutenberg is an online repository of over 33,000 books and pieces of literature with the goal of the creation and distribution to the masses of eBooks ([www.gutenberg.org/wiki/mainpage](http://www.gutenberg.org/wiki/mainpage)). The Project contains all manners of genres and styles of texts, ranging several centuries up to the early 1900s.

The question that this portion attempted to answer was the likelihood that a document supports a statement that violence is unacceptable. From the Project, four variables/classifiers were chosen into which the texts were separated and against which the corpus was tested. The test texts were drawn from a larger collection of texts about which the details (genre, author gender, etc.) were known. The four test classifiers were genre (biography, cookery, horror with 14 texts per genre), gender of author (14 female authors, 14 male), time period written (1700s and earlier, 1800s, 1900s with eight texts in each) and violence (acceptable or unacceptable, with 14 texts each). To test the accuracy of the classifiers and STANLEY’s ability to categorize effectively, all the texts from the

original collection were used as the document corpus (182 texts total). To help ensure solid variable texts for our violence classifiers, all texts for the `_unacceptable` classifier were children's stories, while the `_acceptable` category contained horror, crime, and war texts.

Once the variables were set and the corpus established, STANLEY was run under various configurations to give an idea of how well the tool was performing. The various configurations allow the analyst to specify which categories classifications can be more stringent than others. For violence to be accepted in this case, the text had to have an overall document score for violence of 0.80, a score for `_acceptable` of 0.85, and a delta score from the `_unacceptable` variable option of 0.05. Bayesian Networks were then developed to determine the probabilities of violence being OK, based on the genre, time period, and author gender. Weka, an open source data mining software tool developed by the University of Waikato was used to generate the BN's.

### C. RESULTS

The results for this particular test with STANLEY were mixed. STANLEY did an extremely accurate job of classifying the `_cooking` genre (22 classified out of 22 actual). However, it had more difficulty with correctly classifying the other two genres. Eleven of 21 biographies were correctly classified, and 147 texts were classified as horror, when there were only 25 actual horror texts in the corpus. Two texts were labeled unknown (scores not high enough to meet threshold criteria for genre). The reason for the high horror count is quite obvious. With only three genres from which to choose and classify, the tool categorized each text based on the comparisons between the configuration file and the highest variable score. The horror genre was scored highest in those texts that were categorized as horror when they in fact were some other genre. The biography genre was under classified for a similar reason. There was too much similarity (cosine similarity) in the concept vectors between the actual horror and biography texts. This phenomenon illustrates the need for clearly defined, specific, orthogonal classifiers to successfully utilize the tool.

STANLEY did have quite a bit of difficulty correctly categorizing gender and time period (see Table 8 for specifics on gender and time period). Of the 172 texts for which a gender classification was scored (10 texts failed to meet gender thresholds), only 25 were classified as written by male authors, when in fact, the number of male authors was 114. For female authorship, STANLEY classified 147 when in fact there were only 51. Similarly, for time period written, STANLEY classified 28 as UNKNOWN, 13 of 11 as 1700s and before, 76 of 54 as 1800s, and 65 of 77 as 1900s. There are perhaps several reasons for this, but most likely the main reason is similar to the cause for the misclassification of the horror and biography genres. Again, the classifiers were not specific enough (‘cooking’ was very specific with good results) and there was not a large enough differential in the cosine similarity to distinguish between variables, too much similarity in word.

Table 7. STANLEY results for number of documents and percentages for author gender and time period written.

Category	Reference Male (% total corpus)	Reference Female (% total corpus)	Test Results Male (% total corpus)	Test Results Female (% total corpus)	True Male - male classified as male (% Ref Male)	False Male -- female authored classified as male (% Ref Female)	True Female -- female classified as female (% Ref Female)	False Female -- male authored classified as female (% Ref Male)	?? Male classified unknown (% Ref Male)	?? Female classified unknown (% Ref Female)
Number of Docs (%)	114(62.6%)	51(28.0%)	25(13.7%)	147(80.8%)	13(11.4%)	8(15.7%)	43(84.3%)	95(83.3%)	6(5.2%)	3(5.9%)
Category	Ref 1700s (% total corpus)	Ref 1800s (% total corpus)	Ref 1900s (% total corpus)	Test Results 1700s (% total corpus)	Test Results 1800s (% total corpus)	Test Results 1900s (% total corpus)	True 1700s -- 1700s classified as 1700s (% Ref 1700s)	False 1700s -- 1700s classified as other (% Ref 1800s and 1900s)	?? 1700s classified unknown	
Number of Docs (%)	11(6.04%)	54(29.7%)	77(42.3%)	13(7.1%)	76(41.8%)	65(35.7%)	6(54.5%)	5(3.8%)	0	
Category	True 1800s -- 1800s classified as 1800s (% Ref 1800s)	False 1800s -- 1800s classified as other (% Ref 1700s and 1900s)	?? 1800s classified unknown (% Ref 1800s)	True 1900s -- 1900s classified as 1900s (% Ref 1900s)	False 1900s -- 1900s classified as other (% Ref 1700s and 1800s)	?? 1900s classified unknown (% Ref 1900s)				
Number of Docs (%)	25(46.3%)	21(23.9%)	8(14.8%)	31(40.3%)	27(41.5%)	19(24.7%)				

Note: There were 17 UNKNOWN gender texts and 40 UNKNOWN time period texts in the reference (classifier) corpus.

For the question regarding the acceptability of violence, the test document corpus claimed 150 of 182 acceptable, 22 unacceptable, and 10 unknown (failed to meet threshold criteria). This corresponds closely to the number of documents within the corpus known to be unacceptable of violence with a false positive rate for unacceptable less than 1%. The difference in results is not surprising given the relative strength of the classifiers used and continues to highlight the need for orthogonal classifiers. Figure 51 shows the Bayesian Network generated by Weka and the conditional probabilities of violence being acceptable or unacceptable based on the three other classifiers. The relationship shown in this network is the dependency of violence on the other classifiers. In other words, the acceptability or unacceptability of violence is determined by the probability distributions within the classifiers.

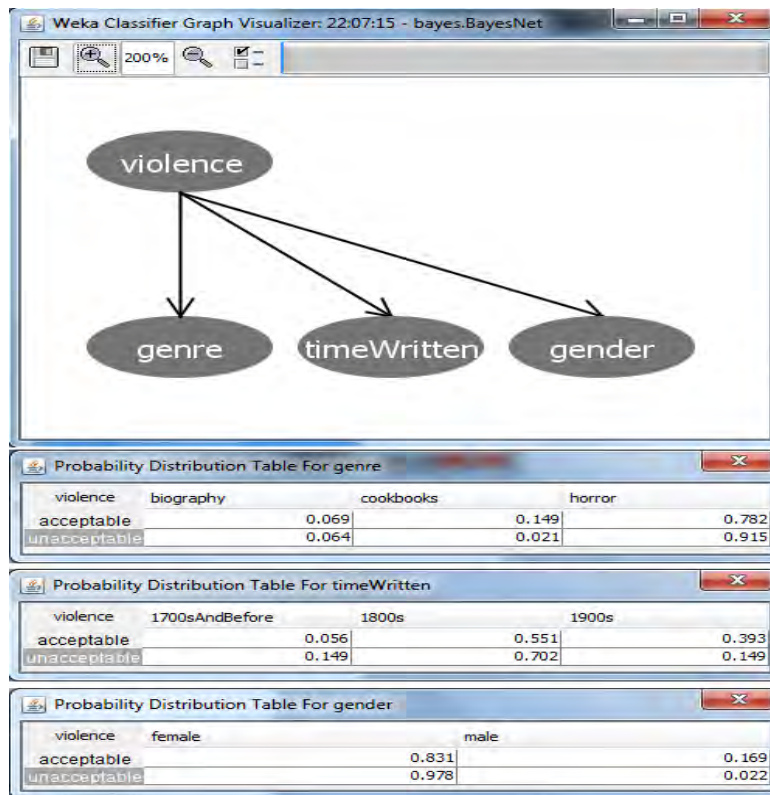


Figure 51. Weka generated Bayesian Network showing the dependency of violence given genre, time written, and author gender

#### **D. LESSONS LEARNED, RECOMMENDATIONS, AND FUTURE WORK**

There are several key conclusions and lessons learned were drawn from this study and applied to subsequent uses of STANLEY in this thesis. The primary lesson learned was that there is potential for a tool like STANLEY in the area of social simulation development. The need for classifiers that are orthogonal cannot be stressed enough. The classifiers need to be as specific as possible to the categories they are attempting to represent. As was shown, when the classifiers were well defined, as was the case with the ‘\_cooking’ genre, STANLEY was very accurate in classifying the documents. When there existed some ambiguity between texts, as could have been the case with the horror and biography genres, the tool was less effective.

There are other problems that can occur with the way that STANLEY classifies documents. This initial implementation using STANLEY was based on classifiers that match based on the statistical use of concepts. This means that a classifier could be more likely to match two documents discussing two different categories using the same style or similar word packets (words, phrases, etc.) than two documents discussing the same subject but using two different styles. To account for different types of documents an ensemble classification method using several different classification technologies in a voting scheme may be more appropriate. The number and type of selected documents used to generate the classification profiles will also be very important and further research is needed to quantify these data needs. Further research is also needed to explore how to best gather and weight the corpus of documents to be classified. It may be that certain types of documents expressing particular points of view may be more prevalent than what is actually observed on the ground due to a media bias or general media access limitations. Some documents may also be deemed more important or credible than others. A mechanism needs to be put in place to resolve these issues.

THIS PAGE INTENTIONALLY LEFT BLANK

## **APPENDIX C. SEED DOCUMENTS FOR SPIDER RUNS**

This appendix includes all the documents used to seed the spider. Particularly, these documents established the classifier list with which the spider scored the documents it retrieved. Included are articles relating to the countries involved in the Middle East Turmoil and articles about the rising oil prices in the U.S. They are included in the format in which they were copied from the Internet (references in section III.B.1)

### **Algeria**

After 19 years, the government officially lifted a state of emergency in February following strikes and protests. But protest marches, which were not allowed under the state of emergency, continue to be banned in the capital, Algiers. Some viewed the move as a “ruse” to placate protesters, who continue to turn out for demonstrations that are quickly broken up by large numbers of police. President Abdelaziz Bouteflika has pledged political reforms. On April 12, thousands of students marched in Algiers to demand the education minister step down and were blocked by police when trying to reach government headquarters.

### **Bahrain**

After a violent crackdown on protesters in the capital, Manama, that killed seven people, the crown prince in February called for a national dialogue between the Sunni-led government and the mostly Shiite protesters. Demonstrators were skeptical of the government’s offer, and they continued to stage daily marches, with many calling for the ouster of the monarchy. Following fighting between protesters and police, a military force from Saudi Arabia and neighboring Gulf states entered Bahrain at the royal family’s request on March 14. A day later, the king declared a three-month state of emergency.

In the crackdown on dissent that followed, security forces cleared demonstrators from the Pearl traffic circle in Manama, imposed a curfew and arrested opposition activists. The government also demolished the monument in the middle of the Pearl roundabout that had become a symbol of the opposition. At least 30 people have been killed since the protests began. On April 14, Bahrain said it was seeking court approval to disband Al Wefaq, the country’s biggest Shiite party, for “threatening peace.”

### **Egypt**

Protesters took to Egypt’s streets in January, demanding the ouster of President Hosni Mubarak. Mubarak supporters clashed with demonstrators in Tahrir Square, which became the focal point of protests in the capital, Cairo. Hundreds were killed in the

uprising. Although Mubarak pledged not to run again, fired his government and appointed a vice president for the first time in his three decades of rule, the protests intensified until Vice President Omar Suleiman announced that the president had handed over power to the military.

Protesters have continued to demand that the military rulers carry out reforms. The military government has said it will lift the country's three-decades-old state of emergency before parliamentary elections scheduled for September. Presidential elections are slated to be held by November. On April 12, Mubarak was hospitalized after he collapsed during questioning by prosecutors looking into corruption allegations and the deaths of protesters during the uprising. Once he was declared stable, the questioning resumed, and a day later, Egypt's prosecutor general announced that Mubarak and his two sons would be detained for 15 days. Mubarak may remain hospitalized for the duration. On April 14, the military government promised to review the cases of young protesters detained in the wake of Mubarak's ouster.

## **Iran**

Tens of thousands of demonstrators turned out Feb. 14 for the biggest protests the country had seen since the aftermath of the disputed re-election of President Mahmoud Ahmadinejad in 2009. After clashes between security forces and the protesters, hard-line lawmakers called for opposition leaders to be put on trial and put to death. On March 1, protesters rallied in Tehran to demand the release of opposition leaders Mir Hossein Mousavi and Mahdi Karroubi, who supporters said had been moved from house arrest to prison. Riot police used tear gas and batons to break up the demonstrations, according to witnesses and opposition websites. Seven people were arrested at the funeral for Mousavi's father on March 31, according to the Iranian reformist website kaleme.com. The country's official IRNA news agency reported April 11 that students threw firebombs at the Saudi embassy to protest Saudi Arabia's involvement in Bahrain.

## **Iraq**

Small, scattered protests, focusing on unemployment, corruption and a lack of services, began taking place in Iraq in early February. Protests intensified in the city of Sulaimaniyah—where demonstrators oppose the leaders of Kurdistan, the semiautonomous region in northern Iraq—and in Basra, where the governor resigned. A nationwide “Day of Rage” called for Feb. 25 turned violent in Mosul and other cities, leading to the deaths of more than a dozen protesters. Protesters have continued to turn out in Baghdad and other cities. On April 11, a leading human rights group criticized the response to the protests, saying the government in Baghdad and the government of the Kurdish region were threatening and harassing journalists and protest organizers.

## **Jordan**

Protesters have been gathering on Fridays to demand more of a voice in government—some want the power to elect their prime minister and Cabinet officials. King Abdullah II fired his Cabinet in February and appointed a new prime minister tasked with carrying out reforms. On March 15, the king set a three-month deadline for agreement on reforms by a committee of government officials and opposition leaders. Hundreds of protesters set up camp in a main square in Amman on March 24, saying they would remain there until the prime minister left and other demands were met. The following day, government supporters clashed with the protesters in the capital. One person died and 120 were injured. On April 12, the leader of an ultraconservative Muslim group said a planned protest had been called off after Jordan released four of its members.

## **Kuwait**

More than 1,000 protesters turned out in Kuwait City on March 8 to call for political changes—including a new prime minister. No violence was reported, but police had blocked off a central square and forced protesters into a parking lot across from a government building. On March 31, the country's official news agency said the Cabinet had resigned over regional turmoil. The move appeared to be an attempt by three ministers to avoid being questioned about why Kuwait did not send troops to Bahrain as part of a Saudi-led force.

## **Libya**

Protests challenging leader Moammar Gadhafi led to a bloody crackdown in February. Amid clashes between opposition forces and troops loyal to Gadhafi, thousands fled Libya, with many crossing borders into Egypt and Tunisia. Rebels quickly took control of much of eastern Libya, with their base in the city of Benghazi, where the anti-Gadhafi uprising began Feb. 15. After weeks of fighting, the regime had consolidated its power in much of the west and was advancing in the east when the U.N. Security Council approved the enforcement of a no-fly zone over Libya on March 17.

An international coalition soon began launching airstrikes and cruise missiles to take out Gadhafi's air defenses and other military targets. Rebels have made gains since the airstrikes began but, lacking in training and equipment, they have been pushed back several times when confronted by Gadhafi's forces. U.S., British and French officials have said arming the anti-government forces is a possibility. Rebels have urged the United States to resume a more central role in the campaign now led by NATO. International groups have warned of a humanitarian crisis in Misrata, the only major western city still under partial rebel control.

## **Morocco**

On Feb. 20, demonstrations were called by a coalition of youth groups, labor unions and human rights organizations demanding greater democracy in the North African kingdom. Several thousand people marched through the capital, Rabat—one of several cities across the country where protests were held. Five people were killed in violence linked to the demonstrations. On March 20, thousands again turned out around the country to press for reforms. King Mohammed VI has announced a plan to revise the country's constitution and says the project will be put to voters in a referendum. The king granted pardons or lightened sentences for 190 prisoners on April 14.

## **Oman**

Protests began in the seaside town of Sohar in late February, resulting in deadly clashes with police. Groups of protesters around the country have since pressed for economic and political reforms. Oman's ruler, Sultan Qaboos bin Said, has ordered 50,000 new jobs and a monthly stipend for the unemployed, and has reshuffled his Cabinet. On March 13, he granted lawmaking powers to officials outside the royal family. One person was killed when protesters clashed with police on April 1.

## **Saudi Arabia**

Police opened fire to disperse a protest March 10 in the eastern city of Qatif. Three protesters and one officer were wounded. Hundreds had gathered to demand the release of political prisoners in a second day of protests in the east, home to the country's Shiite minority. Demonstrations have continued in the east, but wider protests called for in the capital, Riyadh, failed to materialize amid a massive show of police force. Protests are officially banned in the mainly Sunni kingdom. King Abdullah has promised to spend billions of dollars on a benefits package that includes money for home loans, new apartments and payments to government workers, students and the unemployed. The country also plans to hold municipal elections in April after a delay of a year and a half.

## **Syria**

Security forces fired on protesters who had gathered in the southern city of Daraa on March 18, killing five people and fueling mass demonstrations. Angry protesters burned government buildings, and the government fired the governor of Daraa province, whom residents had accused of corruption. The protests have grown steadily, and a violent crackdown across the country has killed at least 200 people, activists said—including 37 killed on April 8. A U.S. State Department spokesman said April 14 that there was "credible information" that Iran was helping Syria crack down on protests.

In response to the push for reforms, the Cabinet resigned in late March, and a new government was formed. President Bashar Assad, who has blamed the dissent on armed

gangs, is facing increasing international criticism over the bloodshed. On April 14, he ordered the release of hundreds of demonstrators who have been detained.

### **Tunisia**

The unrest in this North African nation began in December, apparently after a 26-year-old man committed suicide when police confiscated the fruits and vegetables he was selling. Anger at a lack of employment and at a leadership viewed as corrupt exploded into demonstrations and clashes with police. A United Nations mission says at least 219 were killed in the weeks of protests. President Zine El Abidine Ben Ali fled to Saudi Arabia on Jan. 14. In late February, Mohamed Ghannouchi, who served as prime minister for 11 years, bowed to protesters' demands and resigned after clashes between demonstrators and riot police. The interim president, Fouad Mebazaa, has called for elections July 24 to pick representatives to write a new constitution.

### **Yemen**

Yemen first saw protests in January, with more sustained demonstrations beginning in February. Demonstrators are calling for the ouster of President Ali Abdullah Saleh, who has ruled for more than 30 years. The government intensified its crackdown in March, with police firing on demonstrators and government supporters clashing with crowds. More than 40 people were killed in clashes on March 18.

Saleh's support has crumbled since then, with more than a dozen top military officials—including some from his own tribe—joining the opposition, along with lawmakers, diplomats and governors. Saleh has warned that the country could slide into civil war. The Parliament put in place emergency laws in March that suspend the constitution and give security forces greater powers of arrest and detention. A mediation proposal from neighboring Gulf nations sparked demonstrations across the country on April 12 because it offered Saleh immunity from prosecution. On April 13, one person was killed when gunmen attacked forces loyal to Saleh's rival, Maj. Gen. Ali Mohsen al-Ahmar. At least one other was killed in clashes between protesters and security forces in Aden

### **Libya**

An uprising against Colonel Muammar Gaddafi's rule on 16 February has developed into an armed conflict pitting rebels against government forces and drawing in a Nato-led coalition with a UN mandate to protect civilians.

At least several thousand people have been killed and many more have been injured as the two sides battle for control over territory.

The country's vital coastal cities are now roughly split between pro-Gaddafi forces controlling the capital, Tripoli, and the west, and rebels controlling Benghazi in the east.

Coalition operations have been largely confined to air attacks. These were initially aimed at imposing a no-fly zone but have latterly included strikes on government armoured forces. The UN's Resolution 1973 authorises ~~all~~ necessary measures short of occupation."

Col Gaddafi's government has accepted an African Union peace proposal to end the eight-week conflict, and the plan is now being submitted to the rebels in Benghazi.

In power since 1969, Col Gaddafi is the longest-serving ruler in Africa and the Middle East, and also one of the most autocratic. He and his allies face an International Criminal Court investigation into alleged crimes against humanity.

The UN believes at least 335,000 people have fled Libya since the beginning of the conflict, including at least 200,000 foreign nationals.

## **Morocco**

King Mohammed VI promised "comprehensive constitutional reform" in response to nationwide protests in February but unrest has continued.

Protesters want some of the king's powers to be handed over to a newly elected government.

While there have been deaths, notably when five people died at a bank that was set on fire, security forces appear to have made an effort to avoid violent confrontations.

The main opposition group has warned the "autocracy" will be swept away unless the government responds to the people's demands.

Morocco has been facing severe economic problems. It has announced an increase in state subsidies to try to counter commodity price rises.

Like Jordan, the country is a monarchy with strong support among sections of the public.

## **Algeria**

President Abdelaziz Bouteflika has been under pressure to change the constitution and limit presidential terms after protests that have continued since January.

Strikes, opposition protests and riots have prompted concern among the ruling elite that the country may succumb to popular unrest.

Attempts by protesters to march through the capital, Algiers, have been broken up by huge numbers of riot police.

The trigger for the unrest appears to be mainly economic - in particular sharp increases in the price of food.

Algeria's government has considerable wealth from its oil and gas exports and is trying to tackle social and economic complaints with a huge public spending programme.

Mr. Bouteflika has been president since 1999. He was re-elected for a third term in April 2009 after winning more than 90% of the vote.

The downfall of President Zine al-Abidine Ben Ali, who quit office in January after weeks of protests against poverty and corruption, inspired pro-democracy activists across the Arab world.

He was forced out of power after nearly a quarter of a century.

Public anger had been sparked by the suicide of a young, unemployed man, Mohamed Bouazizi, who set fire to himself on 17 December after officials had blocked his attempts to make a living selling vegetables.

Since the exit of Mr. Ali, who later suffered a stroke, an interim government of technocrats has been set up pending elections on 24 July.

One key reform already enacted is the dissolution of the notorious political police and state security apparatus, which was blamed for many human rights abuses.

## **Egypt**

The military have been running the country since President Hosni Mubarak, in power for three decades, resigned on 11 February following weeks of protests in the capital, Cairo, and other cities.

They are meant to oversee the country's transition to a genuine democracy this summer and Essam Sharaf, a popular former transport minister who sided with the protesters, has been appointed caretaker prime minister.

While much of the country has returned to normal, the protesters have staged weekly Friday demonstrations in Cairo's Tahrir Square to keep pressure on the military to deliver a swift transition to democracy.

The opposition Muslim Brotherhood group is expected to do well in any free and fair election, but fears of a lurch towards Islamist rule is the main worry for Western powers and Israel.

Mr. Mubarak has been summoned by the state prosecutor for questioning over alleged corruption and killings of protesters - accusations which he denies. Mr. Mubarak, his

sons and their wives have been banned from leaving the country and their assets have been frozen.

Much of the unrest was driven by poverty, rising prices, social exclusion and anger at corruption and personal enrichment among the political elite, as well as a demographic bulge of young people unable to find work.

Almost 400 people died and more than 6,000 were injured during the protests.

## **Jordan**

Unrest has been simmering since January but while protesters have clashed with security forces, the country saw nothing like the deadly violence in Syria and Egypt until 25 March, when a man was killed in the capital, Amman.

Protesters have been demanding better employment prospects and cuts in food and fuel costs, as well as electoral reforms that would see the prime minister directly elected and more powers granted to parliament.

King Abdullah II has replaced his prime minister with Marouf al-Bakhit, a former general and ambassador to Israel, together with a new cabinet.

However, a powerful Islamist opposition group, the Islamic Action Front, has called for the dissolution of parliament and criticised the king's efforts to initiate reform.

The Hashemite Kingdom of Jordan is a small country with few natural resources, but it has played a pivotal role in the struggle for power in the Middle East.

## **Syria**

The wave of popular unrest sweeping the Arab world came late to the state run by President Bashar al-Assad, one of the region's most authoritarian rulers.

Rights activists say about 200 Syrians have died since the protests broke out in mid-March at the city of Deraa, before spreading.

After four weeks, demonstrations spread to the cities of Homs and Baniyas, and the government branded them ~~an~~ "an armed insurrection."

Protesters have been demanding greater political freedom, an end to corruption, action on poverty and the end of an emergency law imposed since 1963.

There have been promises of reform in response to the unrest, but Mr. Assad has been making similar pledges to little effect since 2000, when he inherited power from his father Hafez.

Events in Syria, one of Israel's most bitter enemies and a strong ally of Lebanon's Hezbollah militants, could impact on the wider Middle East

## **Saudi Arabia**

The challenge for the rulers of one of the region's wealthiest and most conservative nations has been to address pressure for reform while combating a growing problem of Islamist violence.

The kingdom, home to some of Islam's most sacred places, has seen no mass pro-democracy protests and opposition movements are banned.

However, there have been some small demonstrations by the Shia Muslim minority in solidarity with protesters in Bahrain.

King Abdullah, 82, is regarded in the Arab world as a supporter of wider Arab interests.

If the Saudis have played a role in the "Arab spring" at all, it has perhaps been to support fellow governments under pressure: Saudi soldiers were sent to Bahrain to help shore up the government and it was to Saudi Arabia that Tunisia's ousted leader, Zine al-Abidine Ben Ali, fled in January.

## **Yemen**

Scores of people have been killed in violence between security forces and demonstrators since protests against the rule of President Ali Abdullah Saleh began on 11 February.

Mr. Saleh, in power for more than three decades, has announced he will not seek another term in office, but his current mandate does not expire until 2013 while protesters demand he surrenders power immediately.

After at least 45 people were killed and 270 injured on 18 March, when protesters came under fire in the capital, Sanaa, the president announced a national state of emergency.

A slew of ministers and ambassadors resigned following the 18 March shootings and Mr. Saleh announced he was sacking the entire cabinet.

Both the U.S. and France have condemned the mounting violence and urged the president to allow peaceful protest.

Mr. Saleh has already rejected several calls for him to resign.

Yemen is the Arab world's most impoverished nation and, even before the current protests, it was becoming increasingly chaotic, with both al-Qaeda and separatist challenges to the government's authority.

## **Oman**

Unprecedented protests erupted at the beginning of March, with the deaths of several people.

The Gulf state, with its large and youthful population, had previously been regarded as stable.

It appears the protesters were not demanding radical political change but specific steps such as job creation and controls on food prices, as well as greater power for the semi-elected parliament and checks on corruption.

The oldest independent state in the Arab world, Oman has been ruled by Sultan Qaboos since he seized power from his father, Sultan Said bin Taimur, in 1970.

The oil-rich country is a popular tourist destination and a long-standing ally of the U.S. and UK.

## **Bahrain**

Unrest in the tiny island state, which started on 14 February and has left more than 25 people dead, has been making headlines because of its status as a key U.S. ally, its previous image of stability and its unusual sectarian divide.

Predominantly Shia Muslim protesters have been demanding action to tackle economic hardship, the lack of political freedom and discrimination in jobs in favour of the governing Sunni Muslim minority.

Bahrain, home of the U.S. Fifth Fleet, is unusual because, like Iraq, it is one of the few Arab states with a Shia majority and, as such, is seen by some as vulnerable to influence from Iran.

For weeks, the demonstrators occupied the centre of the capital, Manama.

King Hamad clamped down hard on 16 March, clearing the protesters' camp in a show of force condemned by the UN as -shocking".

He imposed a state of emergency and used hundreds of Saudi and UAE soldiers to beef up his security.

Rights groups say the government has detained more than 400 people - including human rights activists, doctors, bloggers and opposition supporters - since the unrest started. Several Shia activists have complained of being tortured while in custody.

The monarchy has offered talks with representatives of the Shia community, but it rejected key opposition demands such as the sacking of his prime minister, a fellow member of the ruling family.

## **Iran**

Long-simmering unrest over the disputed 2009 presidential election boiled over again on 14 February.

Thousands of people heeded calls by the two main opposition leaders to rally in the capital Tehran in solidarity with pro-democracy protests across the Middle East.

Security forces cracked down on the protest. Two people were killed and many more injured. Further protests on 20 February and 1 March were also suppressed.

Supporters of the government have been calling for the opposition leaders, Mehdi Karoubi and Mir Hossein Mousavi, to be executed. They have been placed under house arrest.

Iran's complex and unusual political system combines elements of a modern Islamic theocracy with democracy. Its President, Mahmoud Ahmadinejad, is a hardliner who has vowed to put down any protests.

Iran's burgeoning nuclear programme has long been regarded with suspicion in Israel and the West, as has its support for militant groups in the Middle East.

Syria—Protests (2011)

Updated: April 20, 2011

Syria's harsh and stagnant dictatorship at first seemed immune to the wave of unrest that swept through most of the Arab world after the revolution in Tunisia in January 2011. But in mid-March, demonstrations broke out in several cities, and grew rapidly after security forces fired on protesters.

The country's president is [Bashar al-Assad](#), the son of Hafez al-Assad, who ruled with an iron hand for three decades before his death in 2000. The Assads belonged to the Allawite sect, a minority that came to hold most of the top positions in the government and military. Under Hafez al-Assad, Syria was reviled in the West for its support of terrorist groups and generally isolated even from more moderate Arab countries. Bashar al-Assad from time to time made gestures toward greater openness. But it remained one of the region's most repressive regimes.

In February 2011, after the fall of Egypt's strongman, Hosni Mubarak, a handful of demonstrations were called in Syria. But the demonstrators were always outnumbered by

the police, and were quickly arrested or dispersed. Large protests began in March in the southern town of Dara'a, where citizens were outraged by the arrest of more than a dozen schoolchildren for writing graffiti. More protests followed, in Dara'a and other parts of the country; after a few days, the regime responded with force. By April 18, human rights groups said they had documented more than 200 deaths of protesters.

The widening defiance in Syria comes after months of challenge to the autocrats of the Arab world that confronts them with a stark choice: either they can step down, as the leaders of Tunisia and Egypt were forced to do, or they can protect their own power with varying degrees of force, like the rulers of Libya, Bahrain and Yemen, at the risk of ever greater violence.

In Syria, the quandary seems particularly acute, with the authorities veering between offers of concessions and crackdowns. On April 16, Mr. Assad pledged to meet one of the demonstrators' main demands by lifting the emergency law. But just four days later, his government threatened to use the provisions of that same law to punish defiance.

### **Protest Timeline**

**April 19** The beleaguered government [bluntly warned its people to end more than a month of demonstrations](#), just hours after it marshaled police, army and other forces to crush one of the biggest gatherings yet by protesters bent on staging an Egyptian-style sit-in in the country's third-largest city. The warning by the Interior Ministry—~~under~~forbidding protests ~~under any banner whatsoever~~—suggested that the government was prepared to escalate a crackdown on dissent. The statement followed another crackdown by government forces on protests, this time in Homs, an industrial city near the Lebanese border, as security forces fired on a crowd of thousands of demonstrators in the city's central square. At the same time, the government announced it would lift a decades-old state of emergency [among other reforms](#) that ostensibly granted civil liberties, curbed the power of the police and abolished draconian courts.

**April 18** More than 10,000 people [occupied a central square in the Syrian city of Homs](#) after funeral processions for some of the 14 people reported killed a day earlier ignited renewed protests.

**April 17** Rejecting the Syrian president's latest effort to mollify them, [thousands of protesters took to the streets in cities and towns](#), using a national holiday commemorating the end of French colonialism to widen their challenge to his family's iron-fisted autocracy. Security officers responded with deadly force, witnesses reported, including live ammunition fired at a funeral and the seizure of critically wounded demonstrators from a hospital.

**April 16** President [Bashar al-Assad](#) of [Syria](#) addressed his nation on in a televised speech aimed at appeasing a two-month-old protest movement that has posed an unprecedented challenge to his family's four decades of rule, according to human rights groups. As he

swore in a new cabinet, Mr. Assad [announced a raft of new legal proposals](#), including a pledge to end the country's 48-year-old emergency law within days, and he expressed sorrow for deaths that have taken place since antigovernment unrest began.

**April 15** Protesters [turned out again in large numbers in cities across Syria](#) to demand reforms, defying a nationwide crackdown in which dozens of demonstrators have been killed by security forces. The marches on Friday were met with tear gas, beatings and reports of gunfire. Seeking to tamp down the unrest, the government of President [Bashar al-Assad](#) had announced several measures that were meant to mollify demonstrators.

**April 13** Syria's [growing protest movement broadened](#) as Aleppo, one of Syria's largest cities, had its first demonstrations against the government of President [Bashar al-Assad](#), and a group of women from the coastal village of Bayda, where hundreds were detained this week, [marched](#) to demand the release of their husbands and sons.

**April 11** Syrian security forces and pro-government gunmen [killed four protesters](#) in the Syrian port city of Baniyas. The army had sealed off the city as hundreds of protesters gathered, undaunted by the government's use of force to quell more than three weeks of unrest, witnesses said. Pro-democracy protests in [Syria](#) spread for the first time to a university campus and were violently suppressed, as [the government made clear](#) there would be ~~no~~ more room for leniency or tolerance."

**April 9** Syrian security forces [fired live ammunition at protesters in two cities](#), a day after the single bloodiest day of [Syria's](#) three-week antigovernment uprising. In Dara, the security forces fired to disperse a funeral march for some of the 37 people killed in protests across the country a day earlier, a human rights group said.

**April 8** [Gunfire erupted after prayers](#) in the southern city of Dara'a as security forces across [Syria](#) moved to counter a third week of protests against the government of President [Bashar al-Assad](#). There were conflicting reports of the number of dead. More than 10,000 protesters took to the streets of several cities, including the capital, Damascus, and a suburb where at least 15 protesters were killed last Friday in clashes with security forces.

**April 6** President [Bashar al-Assad's](#) government offered several unusual gestures intended to earn it good will among Sunnis and Kurds. It announced that Syria's first and only casino, which had enraged Islamists when it opened on [New Year's Eve](#), would be closed. It also said that schoolteachers who had been dismissed in 2010 for wearing the [niqab](#), a type of face veil, would be allowed back to work.

**April 4** The government announced that President [Bashar al-Assad](#) had appointed Adel Safar, the minister of agriculture for the past eight years, [as the new prime minister](#). Meanwhile, thousands of Syrians marched through the shuttered streets of Douma, just outside Damascus, chanting antigovernment slogans as they buried at least eight victims

of the crackdown on protests held April 1. Human rights groups put the death toll from the protests at over 100 and scores of arrests continue.

**April 1** Thousands of protesters [took to the streets](#) in cities around Syria to chants of “We want freedom” and security forces responded with tear gas, electrified batons, clubs and bullets, activists and residents said in telephone interviews. The protests, organized via social networking sites and using Friday prayers as a meeting point, appeared to pose a critical test of the strength of the movement, which in a little more than two weeks has presented an unprecedented challenge to the four-decade iron rule of President [Bashar al-Assad](#) and his family.

**March 31** The government announced that it was creating committees to address the protesters’ concerns but failed to promise immediate action and the move appeared unlikely to quell the rising tide of unrest.

**March 30** In [his first address to the nation](#) after bloody protests and calls for reform, President [Basher al-Assad](#) blamed a broad conspiracy from beyond his borders for [Syria’s](#) turmoil and offered no concessions to ease his authoritarian regime’s grip on public life. He acknowledged that “Syrian people have demands that have not been met,” but said that those grievances were “used as a cover to dupe the people to go to the streets.” He added that “some of them had good intentions.”

**March 29** President [Bashar al-Assad](#) accepted the [resignation of his cabinet](#) as tens of thousands of government supporters took to the streets of the capital in an effort to counter a rising tide of pro-democracy protests in several cities. The cabinet resignation marked a rare moment of responsiveness to public pressure by the Syrian government, which has taken a carrot-and-stick approach to a deepening political crisis.

**March 28** Syrian forces fired into the air to disperse hundreds of protesters in Dara’a calling for an end to emergency laws, but demonstrators regrouped despite a heavy troop deployment, a witness said.

**March 26** President [Bashar al-Assad](#) of the ruling [Baath Party](#) began the day in what appeared to be a gesture intended to ease the crisis, when he announced the release of as many as 200 political prisoners. But by sunset, Baath Party offices were burning in at least two cities, the military was deployed in Latakia and once again government forces opened fire with live rounds, witnesses said. Human rights groups put the confirmed death toll in protests so far at 61.

**March 25** Military troops [opened fire during protests](#) in the southern part of [Syria](#) and killed peaceful demonstrators, according to witnesses and news reports, hurtling the strategically important nation into turmoil. Tens of thousands took to the streets in protest around the nation, defying a state that has once again demonstrated its willingness to use lethal force.

**March 23** Security forces began a crackdown in Dara'a, after the Syrian Army reinforced the police presence and confronted a group of protesters who had gathered in and around the Omari mosque in the city center. Mr. Assad promised increased freedoms for discontented citizens and increased pay and benefits for state workers. High-ranking aides said that the army would not shoot peaceful demonstrators and spoke of lifting the 50-year-old state of emergency.

**March 21** Demonstrators in Dara'a set fire to the ruling [Baath Party's](#) headquarters and other government buildings. Police officers [fired live ammunition into the crowds](#), killing at least one and wounding scores of others, witnesses said. Mr. Assad made some conciliatory gestures, but crowds continued to gather in and around the Omari mosque in Dara'a, chanting their demands: the release of all political prisoners; trials for those who shot and killed protesters; the abolition of Syria's 48-year emergency law; more freedoms; and an end to pervasive corruption.

### **Background to Protests**

The country's last serious stirrings of public discontent had come in 1982, when increasingly violent skirmishes with the Muslim Brotherhood prompted Hafez al-Assad to move against them, sending troops to kill at least 10,000 people and smashing the old city of Hama. Hundreds of fundamentalist leaders were jailed, many never seen alive again.

Syria has a liability not found in the successful uprisings in Tunisia and Egypt—it is a majority Sunni nation that is ruled by a religious minority, the Alawite sect of Shiite Islam. Hafez Assad forged his power base through fear, cooption and sect loyalty. He built an alliance with an elite Sunni business community, and created multiple security services staffed primarily by Alawites. Those security forces have a great deal to lose if the government falls, experts said, because they are part of a widely despised minority, and so have the incentive of self-preservation.

### **Foreign Policy**

Under the administration of President George W. Bush, Syria was once again vilified as a dangerous pariah. It was linked to the 2005 killing of a former Lebanese prime minister, [Rafik Hariri](#). In 2007, Israeli jets destroyed buildings in Syria that intelligence officials said might have been the first stage in a nuclear weapons program. And the United States and its Arab allies mounted a vigorous campaign to isolate Damascus, which they accused of sowing chaos and violence throughout the middle east through its support for militant groups like [Hezbollah](#) and [Hammas](#).

President Obama came into office pledging to engage with Syria, arguing that the Bush administration's efforts to isolate Syria had done nothing to wean it from Iran or encourage Middle East peace efforts. So far, however, the engagement has been limited. American diplomats have visited Damascus, but have reiterated the same priorities as the

Bush administration: protesting Syria's military support to Hezbollah and [Hamas](#), and its strong ties with Iran.

Secret State Department cables obtained by [WikiLeaks](#) and made available to several news organizations show that arms transactions involving Syria and Hezbollah continue to greatly concern the Obama administration. Hezbollah's arsenal now includes up to 50,000 rockets and missiles, including some 40 to 50 Fateh-110 missiles capable of reaching Tel Aviv and most of Israel, and 10 Scud-D missiles.

—Syria's determined support of Hizballah's military build-up, particularly the steady supply of longer-range rockets and the introduction of guided missiles could change the military balance and produce a scenario significantly more destructive than the July-August 2006 war," said a November 2009 cable from the American chargé d'affaires in Damascus.

According to cables, Syrian leaders appeared to believe that the weapons shipments increased their political leverage with the Israelis. But they made Lebanon even more of a tinderbox and increased the prospect that a future conflict might include Syria.

### **The Hariri Case**

Also looming is potential new trouble in Lebanon, where a [United Nations](#)-backed international tribunal is expected to indict members of Hezbollah in the death of Mr. Hariri. Hezbollah and its allies—including high-ranking Syrian officials—have warned that an indictment could set off civil conflict.

The United States withdrew its ambassador in 2005 after Mr. Hariri was killed in a car bombing in Beirut along with 22 others. Syria was widely accused of having orchestrated the killing, though it has vehemently denied involvement. The Bush administration imposed economic sanctions on Syria, as part of a broader effort to isolate the government of President [Bashar al-Assad](#).

The current chill is a significant change from the situation a few years ago, when Mr. al-Assad showed signs of wanting warmer relations with the West than his father, Hafez al-Assad, had ever pursued. President [Nicolas Sarkozy](#) of France led the way with a visit in September 2008. [King Abdullah of Saudi Arabia](#), who was said to be furious at the Syrian president, [Bashar al-Assad](#), welcomed him warmly in the Saudi capital, Riyadh, in March 2009. And Prime Minister Ehud Olmert of Israel hinted at a revival of talks on the Golan Heights -- a prospect that faded when Mr. Olmert was succeeded by the more conservative Benjamin Netanyahu.

Libya, an oil-rich nation in North Africa, has been under the firm, if sometimes erratic, leadership of Col. [Muammar el-Qaddafi](#) since he seized power in 1969. But in February 2011, the unrest sweeping through much of the Arab world [erupted in several Libyan cities](#). Though it began with a relatively organized core of antigovernment opponents in

Benghazi, its spread to the capital of Tripoli was swift and spontaneous. Colonel Qaddafi lashed out with a level of violence unseen in either of the other uprisings, but an inchoate opposition cobbled together the semblance of a transitional government, fielded a makeshift rebel army and portrayed itself to the West and Libyans as an alternative to Colonel Qaddafi's erratic control.

Momentum shifted quickly, however, and the rebels faced the possibility of being outgunned and outnumbered in what increasingly looked like a mismatched civil war. As Colonel Qaddafi's troops advanced to within 100 miles of Benghazi, the rebel stronghold in the west, the [United Nations Security Council](#) voted to authorize military action, a risky foreign intervention aimed at averting a bloody rout of the rebels by loyalist forces. On March 19, American and European forces began a broad campaign of strikes against Colonel Qaddafi and his government, unleashing warplanes and missiles in a military intervention on a scale not seen in the Arab world since the Iraq war.

The attacks prompted two of Colonel Qaddafi's sons to float a proposal that would remove him from power, which the rebels rejected. Meanwhile, their ragtag forces surged forward and back, unable to make progress against the army despite the help from above but no longer in grave peril. In mid-April, Britain, France and Italy said they would send military liaison officers to help the rebels, a tacit admission that the airstrikes had failed to disable the government's forces.

#### **LATEST DEVELOPMENTS:**

**April 20** The French and Italian governments said that they would join Britain in [sending a small number of military liaison officers](#) to support the ragtag rebel army in [Libya](#), offering a diplomatic boost for the insurgent leader, Mustafa Abdel-Jalil, as he met with President [Nicolas Sarkozy](#) in Paris.

**April 19** [Britain](#) will [send experienced military officers](#) to [Libya](#) to advise rebels fighting forces loyal to Col. [Muammar el-Qaddafi](#). The soldiers marching orders are to help the rebels' makeshift force ~~improve~~ their military organizational structures, communications and logistics," Britain's foreign secretary, [William Hague](#), said in a statement.

**April 15** Forces loyal to Col. [Muammar el-Qaddafi](#), surrounding Misurata and vowing to crush the rebellion there, [fired into residential neighborhoods](#) with heavy weapons, including [cluster bombs](#), which were banned by much of the world. And divisions in [NATO](#) over the intensity of the air campaign [emerged again](#) on the second day of a gathering of foreign ministers. While the British foreign secretary said efforts to persuade more countries to adopt a tougher military posture toward the forces of Colonel Qaddafi were ~~making~~ a bit of progress," the Italian defense minister said his country would not order its pilots to open fire over [Libya](#).

**April 14** NATO foreign ministers gathered to wrestle with increasingly complex [questions raised by the stalled conflict in Libya](#), seeking a formula for political progress

in the absence of any decisive military gains. Pentagon officials disclosed that American warplanes had [continued to strike targets](#) there even after the Obama administration said the United States was stepping back from offensive missions and letting [NATO](#) take the lead.

**April 13** [NATO](#), Arab and African ministers met with [Libya's](#) rebels here in a [show of support for insurgents](#) who are seeking to overthrow Col. [Muammar el-Qaddafi](#) against a backdrop of division over the pace of coalition air attacks on pro-Qaddafi forces. [France](#) and [Britain](#) have openly called on the alliance and its partners to intensify airstrikes on Libyan government troops to protect civilians, prompting an unusual public retort from NATO's command.

**April 11** [African Union](#) negotiators faced a chilly reception upon arriving in eastern [Libya](#) to try [brokering a cease-fire](#) with Libyan rebels, a day after Col. [Muammar el-Qaddafi's](#) military forces appeared to falter in their assault against the rebel side in the battle for the strategic city of Ajdabiya.

**April 9** Military forces loyal to Col. [Muammar el-Qaddafi](#) pressed a [coordinated ground attack on Ajdabiya](#), bringing the front lines of the battle with Libyan opposition forces back to the doorstep of this strategically vital rebel city.

**April 6** Stung by criticism from rebel leaders, [NATO](#) officials said that [the pace of attacks](#) on the forces of Col. [Muammar el-Qaddafi](#) was increasing, after a slight slowdown as the coalition handed off responsibility earlier in the week. Gen. Abdul Fattah Younes, the head of the rebel army, had lashed out at his Western allies during a news conference in Benghazi, accusing NATO of tardiness and indecision.

**April 5** Forces loyal to Col. [Muammar el-Qaddafi](#) battered rebel fighters on the road outside the strategic [oil](#) town of Brega with rocket fire, mortars and artillery, [driving them many miles to the north and leaving them in disarray](#). Colonel Qaddafi's son, [Seif al-Islam](#), promised in a television interview to usher in a new era of constitutional democracy in which his father would be a mere figurehead —like the queen of England.”

**April 4** The United States [began to remove](#) its warplanes from front-line missions in Libya and focus on a support role there. The changeover came as diplomatic maneuvering quickened with Turkey announcing efforts to secure a cease-fire and Italy saying it was recognizing the rebels seeking to oust Col. [Muammar el-Qaddafi](#), only the third country to do so. The Obama administration also dropped financial sanctions against [Moussa Koussa](#), the top Libyan official who fled to [Britain](#), saying it hoped the move would encourage other senior aides to abandon Col. [Muammar el-Qaddafi](#), the country's embattled leader.

**April 3** At least two sons of Col. [Muammar el-Qaddafi](#) are proposing a resolution to the Libyan conflict [that would entail pushing their father aside](#) to make way for a transition to a constitutional democracy under the direction of his son [Seif al-Islam el-Qaddafi](#). At

the same time, as the struggle with Colonel Qaddafi threatened to settle into a stalemate, the rebel government here was [showing growing strains](#) that imperil its struggle to complete a revolution and jeopardize requests for foreign military aid and recognition.

**April 1** A senior aide to one of Col [Muammar el-Qaddafi](#)'s sons [held secret talks in London with British authorities](#), adding to the confusion swirling around the Tripoli regime. East of Brega, the Libyan rebels prepared for a further attempt to wrest the momentum of ground fighting away from Colonel Qaddafi's forces after days of see-sawing advances and retreats. In Washington, [President Obama](#)'s top two national security officials signaled that [the United States was unlikely to arm the rebels](#). Members of the [NATO](#) alliance said they had sternly [warned the rebels](#) not to attack civilians.  
[Timeline: Qaddafi](#)

**March 31** Col. Muammar el-Qaddafi's forces [pushed rebels into a panicked retreat](#) and seized valuable oil towns they ceded just days ago under allied airstrikes. [Libya](#)'s foreign minister, Moussa Koussa, defected to London, dealing a blow to Colonel Qaddafi's government even as his forces made military advances. American officials revealed that the [Central Intelligence Agency](#) has [inserted clandestine operatives into Libya](#) to gather intelligence for military airstrikes and to contact and vet the beleaguered rebels.

**March 30** Leaders of four dozen countries meeting in London agreed that Col. Muammar el-Qaddafi would have to [relinquish power](#), even though regime change is not the stated aim of the [United Nations](#) resolution authorizing military action against his forces. With the [momentum of ground combat tilting in favor of forces loyal to Colonel Qaddafi](#), rebels seeking to oust him embarked on a large-scale withdrawal from the coastal oil town of Brega, falling back toward the strategically located city of Ajdabiya. The Obama administration engaged in a fierce debate over whether to supply weapons to the rebels, with some fearful that providing arms would deepen American involvement in a civil war and that some fighters may have links to [Al Qaeda](#).

**March 29** In his first major address since ordering American airstrikes, [President Obama](#) defended the American-led military assault in Libya, saying it was [in the national interest of the United States to stop a potential massacre](#) and that the assault would be limited. An array of diplomats and public figures gathered in London to shape their political vision of a post-Qaddafi era. In Libya, rebels seeking the ouster of Colonel Qaddafi traded rocket fire with loyalist forces, [who have blunted the insurgents' westward advance](#). At the same time, American warplanes appeared to have opened a new line of attack on pro-Qaddafi forces, firing on three Libyan vessels off the contested western port of Misurata.

## BACKGROUND

Colonel Qaddafi took power in a bloodless coup in September 1969 and has ruled with an iron fist, seeking to spread Libya's influence in Africa. He has built his rule on a cult of

personality and a network of family and tribal alliances supported by largess from Libya's oil revenues.

The United States withdrew its ambassador from Libya in 1972 after Colonel Qaddafi renounced agreements with the West and repeatedly inveighed against the United States in speeches and public statements.

After a mob sacked and burned the American Embassy in 1979, the United States cut off relations. But the relationship did not reach its nadir until 1986, when the Reagan administration accused Libya of ordering the bombing of a German discothèque that killed three people. In response, the United States bombed targets in Tripoli and Benghazi.

The most notorious of Libya's actions was the bombing in 1988 of Pan Am Flight 103 over Lockerbie, Scotland, which killed 270 people. Libya later accepted responsibility, turned over suspects and paid families of victims more than \$2 billion.

After a surprise decision to renounce terrorism in 2003, Colonel Qaddafi re-established diplomatic and economic ties throughout Europe. He had also changed with regard to Israel. The man who once called for pushing the Zionists into the sea advocated the forming of one nation where Jews and Palestinians would live together in peace.

Rather than trying to destabilize his Arab neighbors, he founded a pan-African confederation modeled along the lines of the European Union. On Feb. 2, 2009, Colonel Qaddafi was named chairman of the African Union. His election, however, caused some unease among some of the group's 53-member nations as well as among diplomats and analysts. The colonel, who has ruled Libya with an iron hand, was a stark change from the succession of recent leaders from democratic countries like Tanzania, Ghana and Nigeria.

The most significant changes had been the overtures Colonel Qaddafi has made toward the United States. He was among the first Arab leaders to denounce the Sept. 11 attacks, and he lent tacit approval to the American-led invasion of Afghanistan. To the astonishment of other Arab leaders, he reportedly shared his intelligence files on Al Qaeda with the United States to aid in the hunt for its international operatives. He had also cooperated with the United States and Europe on nuclear weapons, terrorism and immigration issues.

In August 2009, Colonel Qaddafi embarrassed the British government and drew criticism from President Obama with his triumphant reaction to the release from prison of [Abdel Basset Ali al-Megrahi](#), the only person convicted in the bombing of Pan Am Flight 103. Mr. Megrahi was given a hero's welcome when he arrived in Libya, and Colonel Qaddafi thanked British and Scottish officials for releasing Mr. Megrahi at a time that they were trying to distance themselves from the action.

Colonel Qaddafi, born in 1942, is the father of many sons who are now jockeying to succeed him. Experts say his eldest, Seif al-Islam el-Qaddafi, is the current leader. Educated in Britain, well-dressed and fluent in English, he has been a bridge between the Libya power centers and the West.

Prior to the 2011 unrest, the only hint of potential change in Libya came from Seif al-Islam el-Qaddafi, [who spoke of dismantling](#) a legacy of Socialism and [authoritarianism](#) introduced by his father 40 years ago. Seif Qaddafi proposed [far-reaching ideas](#): tax-free investment zones, a tax haven for foreigners, the abolition of visa requirements and the development of luxury hotels.

Seif Qaddafi liked to boast that his country could be “the Dubai of North Africa,” he said, citing Libya’s proximity to Europe (the flight from London to Tripoli is under three hours), its abundant energy reserves and 1,200 miles of mostly unspoiled Mediterranean coastline. Libya is wealthier than debt-ridden, oil-poor Dubai. Its \$15,000 gross domestic product per person ranks it above Poland, Mexico and Chile, according to the [World Bank](#). The government’s sovereign fund, a reserve of oil revenues, boasts \$65 billion. And the government has announced plans to invest \$130 billion over the next three years to improve infrastructure.

But the reality of daily life in Tripoli remained far removed from those lofty notions. The streets were strewn with garbage, there were gaping holes in the sidewalks, tourist-friendly hotels and restaurants were few and far between. And while a number of seaside hotels were being built, the city largely ignored its most spectacular asset, the Mediterranean.

Unemployment is estimated as high as 30 percent and much of the potential work force is insufficiently trained.

### **Uprising in Libya**

In February 2011, protests broke out in several parts of Libya on a so-called [Day of Rage to challenge Colonel Qaddafi’s 41-year-old iron rule](#)—the region’s longest. Thousands turned out in the restive city of Benghazi; in Tripoli; and at three other locations, according to [Human Rights Watch](#). The state media, though, showed Libyans waving green flags and shouting in support of Colonel Qaddafi.

Trying to demonstrate that he was still in control, Colonel Qaddafi appeared on television on Feb. 22, 2011, speaking from his residence on the grounds of an army barracks in Tripoli that still showed scars from when the United States bombed it in 1986.

Colonel Qaddafi, who took power in a military coup, has always kept the Libyan military too weak and divided to rebel against him. About half of Libya’s relatively small 50,000-member army is made up of poorly trained and unreliable conscripts, according to the [Center for Strategic and International Studies](#).

Many of its battalions are organized along tribal lines, ensuring their loyalty to their own clan rather than to top military commanders—a pattern evident in the defection of portions of the army to help protesters take the eastern city of Benghazi. Some Libyans and scholars outside the country say this system of tribal alliances, long Colonel Qaddafi's most potent weapon, is now emerging as perhaps a potential vulnerability.

His own clan dominates the air force and the upper level of army officers, and they are believed to have remained loyal to him, in part because his clan has the most to lose from his ouster.

Distrustful of his own generals, he built up an elaborate paramilitary force—accompanied by special segments of the regular army that report primarily to his family. It is designed to check the army and in part to subdue his own population. At the top of that structure is his roughly 3,000-member revolutionary guard corps, which mainly guards him personally.

But perhaps the most significant force that Colonel Qaddafi has deployed against the current insurrection is one believed to consist of about 2,500 ruthless mercenaries from countries like Chad, Sudan and Niger that he calls his Islamic Pan African Brigade.

### **The Ongoing Conflict**

On Feb. 25, security forces loyal to Colonel Qaddafi used gunfire to try to disperse thousands of protesters [who streamed out of mosques after prayers](#) to mount their first major challenge to the government's crackdown in Tripoli. Rebel leaders said they were sending forces from nearby cities and other parts of the country to join the fight.

The ring of rebel control around Tripoli tightened, but in a sign that the fight was far from over, armed government forces massed around the city.

The [United Nations Security Council](#) voted unanimously to impose sanctions on Colonel Qaddafi and his inner circle of advisers, and [called for an international war crimes investigation](#) into “widespread and systemic attacks” against Libyan citizens.

On March 2, rebels in the strategic [oil](#) city of Brega repelled an attack by hundreds of Colonel Qaddafi's fighters. The daylong battle was the first major incursion by the colonel's forces in the rebel-held east of the country since the Libyan uprising began.

Air power proved to be Colonel Qaddafi's biggest advantage, and rebels were unable to use bases and planes they captured in the east. Planes and helicopters gave the Qaddafi forces an additional advantage in moving ammunition and supplies, a crucial factor given the length of the Libyan coast between the rebel stronghold of Benghazi and Tripoli.

As Colonel Qaddafi's forces tried to retake a series of strategic oil towns on the east coast of the country, which fell early in the rebellion to antigovernment rebels, the West continued to debate what actions to take.

### **Western Involvement**

After days of often acrimonious debate played out against a desperate clock, [the Security Council authorized member nations to take –all necessary measures” to protect civilians](#), diplomatic code words calling for military action. Benghazi erupted in celebration at news of the resolution's passage.

A military campaign against Colonel Qaddafi, under British and French leadership, was launched less than 48 hours later. American forces mounted a campaign to knock out Libya's air defense systems, firing volley after volley of Tomahawk missiles from nearby ships against missile, radar and communications centers. Within a week allied air strikes had averted a rout by Colonel Qaddafi of Benghazi and established a no-fly zone over Libya.

The campaign, however, was dogged by friction over who should command the operation, with the United States eventually handing off its lead role to [NATO](#), and by uncertainty over its ultimate goal. Western leaders acknowledged that there was no endgame beyond the immediate United Nations authorization to protect Libyan civilians, and it was uncertain whether even military strikes would force Colonel Qaddafi from power.

In a nationally televised speech March 28, [President Obama](#) defended the American-led military assault, emphasizing that it would be limited and insisting that America had the responsibility and the international backing to stop what he characterized as a looming genocide. At the same time, he said, directing American troops to forcibly remove Colonel Qaddafi from power would be a step too far, and would –splinter” the international coalition that has moved against the Libyan government.

The question of the opposition's capabilities is likely to prove decisive to the fate of the rebellion, which no longer appears outmatched by government forces or troubled by tribal divisions that the government sought to exploit.

But as they seek to capitalize on the damage from Western airstrikes, rebel forces in Libya are fired more by enthusiasm than experience. The political leadership has virtually begged the international community to recognize it, but it has yet to impose its authority in regions it nominally controls.

Meanwhile, the American military warned that the insurgents' rapid advances could quickly be reversed without continued coalition air support.

## Yemen—Protests (2011)

Yemen is a poor, deeply divided country that has been in turmoil since January 2011, when the example of the Tunisian revolution set off mass demonstrations against President [Ali Abdullah Saleh](#). Mr. Saleh, who has been in power since 1978, responded alternately with conciliatory measures, including a promise including an offer not to seek reelection, and violent crackdowns,

As demonstrations continued, Mr. Saleh's support began to crumble, as some army commanders and tribal leaders called for his ouster. The United States, which had long supported Mr. Saleh, even in the face of the protests, [quietly shifted positions](#) after concluding that he is unlikely to bring about the required reforms and must be eased out of office. On April 7, an organization of oil-rich Persian Gulf states [joined the increasing number of international voices calling for a transfer of presidential powers](#) in which Mr. Saleh would hand power to a government of national unity.

The turning point appears to have come on March 18 in a bloody but failed attempt to break the back of the protest. As tens of thousands of demonstrators raised from their noon prayers, security forces and government supporters opened fire. At least 50 people were killed and more than 100 injured, dwarfing the level of violence in previous clashes, but it failed to disperse the crowd. Mr. Saleh declared a state of emergency shortly after the violence, and denied that security forces had been involved in any shooting.

As the demonstrations continued, Mr. Saleh fired his cabinet. On March 21, five army commanders and one of the country's most important tribal leaders threw their support behind the protesters, calling for Mr. Saleh's immediate ouster. A stream of Yemeni officials resigned from the government, including the mayor of the restive southern city of Aden, a provincial governor and at least one of the country's ambassadors.

Yemen's opposition coalition, the Joint Meetings Parties, proposed a plan under which Mr. Saleh would leave at the end of 2011, and he agreed. But protesters then rejected the plan and called for Mr. Saleh's immediate ouster. Opposition leaders said they would travel to Riyadh, the Saudi capital, at the end of April to meet with [Gulf Cooperation Council](#) officials about an initiative discussed earlier in the month. It calls for Mr. Saleh to transfer presidential powers to his deputy and leave office. [The draft agreement](#) also gives the president and his family immunity from prosecution, presumably to head off a situation similar to that in Egypt, where the military [has detained](#) former President [Hosni Mubarak](#) and his two sons.

[Meanwhile, protests and deadly clashes with security forces continue](#), and [rival military factions allied with the government and the rebels](#) fought in Sana on April 13. More than 100 protesters have died since the turmoil began.

## Overview

Home to one of the world's oldest civilizations, Yemen is the poorest country in the Arab world as well as a haven for Islamic jihadists and the site of what amounts to a secret American war against leaders of a branch that [Al Qaeda](#) has established there.

Until the protests, the world's attention had mainly been focused on fears that the country could become [Al Qaeda's](#) next operational and training hub, rivaling the lawless tribal areas of Pakistan. Yemen's stability was of increasing concern to the United States, which has provided \$250 million in military aid in the past five years. The Obama administration was nurturing enduring ties with Mr. Saleh's government to prod him to combat Al Qaeda. The U.S. military was conducting airstrikes even before the Christmas Day 2009 attempted bombing of a Detroit-bound jet by a [23-year-old Nigerian man](#) who later claimed that Qaeda leaders in Yemen had trained and equipped him.

The delicate position of the United States in dealing with Mr. Saleh now seems as evident in Yemen as it is in [Bahrain](#), where pro-American leaders have cracked down on adversaries on the street clamoring for the monarchy to make way for democratic change.

Diplomatic cables obtained by [WikiLeaks](#) and made available to several news organizations offered [an intimate view](#) of the wily, irreverent and sometimes erratic Yemeni autocrat. Mr. Saleh has sometimes accommodated and other times rebuffed American requests on counter terrorism.

## History

With its location at the southwestern end of the Arabian Peninsula, the land of ancient Yemen became rich from the spice trade. So rich that the Romans called the land Arabia Felix—Happy Arabia—and Augustus Caesar tried, but failed, to annex it. That prosperity overlapped with the rule of an Islamic caliphate in the 7th century. When the caliphate broke up, Islamic imams exerted control, sowing the seeds of a theocratic political system that would survive for centuries.

Northern Yemen became part of the Ottoman Empire. Southern Yemen was in the hands of the British after 1839, when they built a protectorate around their port of Aden. North Yemen would become independent of the Ottoman Empire in 1918 and declare itself a republic in 1962; it was not until 1967 that the British withdrew from southern Yemen.

When Marxists took over the government of southern Yemen in 1970, many people fled to the north, and a civil war raged for two decades. The conflict became a proxy conflict in the cold war, with the Soviet Union aiding South Yemen, and the United States bolstering the north.

Though north and south were unified as the Republic of Yemen on May 22, 1990, the violence and internecine conflict did not end. The country's extreme topography—with

dramatically rugged mountains and remote deserts—helped create impenetrable fortresses for warring tribes, which have long attacked government officials and foreign tourists, as well as one another.

Today Yemen faces a violent separatist movement in the south and an intermittent [rebellion in the northwest](#), though President Saleh has expertly played Yemen's various tribes and factions against one another for decades. When one of the country's most prominent tribal sheiks, Hussein al-Ahmar, [resigned from the ruling party](#), it was a deeply troubling sign for the regime.

### **Al Qaeda in Yemen**

Much of the violent tribal feuds, banditry and kidnapping in Yemen appear beyond the control of the central government. Yemen has the region's largest arms market: the country, with roughly 20 million people, is said to have at least 20 million guns.

Yemen did not become a special concern for the United States until 2000, when Qaeda operatives blasted a hole in the American destroyer Cole, killing 17 sailors. After the terrorist attacks of Sept. 11, Yemen joined in a counterterrorism partnership with the United States, and its American-trained forces had some successes in fighting jihadists, even as terrorist attacks on foreign targets continued sporadically.

The jihadists claiming allegiance to Al Qaeda appear to have reorganized and become more methodical, releasing more propaganda materials on the Internet and carrying out more attacks. In July 2007, suicide bombers killed seven Spanish tourists in eastern Yemen, and there were two unsuccessful attacks on oil installations. In September 2008, 10 people were killed (none of them Americans) when two car bombs were detonated outside the American embassy in Sana, the capital.

[Anwar al-Awlaki](#), the American-born cleric now hiding in Yemen, is perhaps the most sophisticated ideological opponent the United States has faced since 2001. Several former [Guantánamo](#) detainees fled in 2009 to Yemen from Saudi Arabia and pledged to mount attacks on Saudi Arabia and other countries from their Yemeni redoubt.

Despite the American airstrike campaign, the leadership of Al Qaeda in the Arabian Peninsula survives, and there is little sign the group is much weaker. Attacks by Qaeda militants in Yemen have conducted several deadly assaults on Yemeni army convoys. Al Qaeda's Yemen branch regularly puts out its first English-language online magazine, Inspire, complete with bomb-making instructions.

### **An Uncertain Future**

Whether or not Mr. Saleh is forced from power, the political crisis in Yemen will likely remain acute, not only because of its tribal culture and topography, but also because of its

deep poverty, high illiteracy and birth rates, and deeply entrenched government corruption. Its economy is precariously tied to oil resources, which are declining rapidly.

The governing elite mainly come from the Sunni majority, which makes up 55 percent of the population and is concentrated in the more developed coastal regions of the south and southwest. A Shiite movement, based in the mountainous north, declared independence and its intermittent rebellion has left thousands of people dead since it began in 2004.

The government is also deeply unpopular in the remote provinces where Al Qaeda militants have sought sanctuary. The tribes there tend to regularly switch sides, making it difficult to depend on them for information about Al Qaeda. —My state is anyone who fills my pocket with money,” goes one old tribal motto.

The current democracy protesters may mark a change from that mindset. During February’s protests in Taiz, long been a bastion of opposition sentiment, a local cleric preached to the crowds of men and women sitting on the pavement.

—This is not a revolution against a person, a family or a tribe,” he said over a loudspeaker to the gathering, which stretched over blocks and blocks of the city’s streets. —This is a revolution against oppression and corruption.”

For years, Mr. Saleh managed tribal-dominated Yemen by propping up scores of carefully chosen tribal leaders, giving them money and weapons and placing them in important positions in government. The loyalty of these empowered sheiks largely guaranteed the loyalty of their followers.

But tribesmen from rural areas made up the majority of the tens of thousands spending day and night at the demonstration in Sana. With large numbers of them unemployed, their vow to stay at Sana’s encampment until Mr. Saleh steps down carried weight.

## **Tunisia**

Until January 2011 Tunisia was known mostly as the most European country of North Africa, with a relatively large middle class, liberal social norms, broad gender equality and welcoming Mediterranean beaches. But now it has taken center stage as the launching pad of a wave of revolt that has swept through the Arab world and beyond.

For all its modern traits, Tunisia had one of the most repressive governments in a region full of police states, and levels of corruption among its elite that became intolerable once the economic malaise that has gripped southern Europe spread to the country.

In what became known as the Jasmine Revolution, a sudden and explosive wave of street protests ousted the authoritarian president, [Zine el-Abidine Ben Ali](#), who had ruled with an iron hand for 23 years. On January 14, Mr. Ben Ali [left the country](#), after trying unsuccessfully to placate the demonstrators with promises of elections. According to

government figures issued later, 78 protesters died and 94 were injured during the demonstrations.

### **Interim Government and Continued Protest**

The prime minister, [Mohamed Ghannouchi](#), created a government of unity, bringing in members of the official opposition, to serve as an interim government until elections could be held in mid-year. But turmoil continued, with new rounds of protests, streams of refugees leaving Tunisia for Europe or entering from Libya. On Feb. 27, [Mr. Ghannouchi resigned](#) in response to complaints that he was too closely tied to Mr. Ben Ali.

Some protesters have called for the complete eradication of the old ruling party, while complaining that outlawed parties like the once powerful Islamist groups or the Tunisian Communists—battle-scarred stalwarts of the long dissident fight against Mr. Ben Ali’s 23-year-rule—were still barred from participating. That movement’s potential reincarnation is perhaps the most significant variable in Tunisia’s post-revolutionary future—yearned for by legions of working-class and rural Tunisians, viewed with just as much apprehension by the cosmopolitan coastal elite.

The revolution, as Tunisians call it, also has created a power vacuum, and Tunisia faces enormous challenges in rebuilding its political system. The country’s caretaker government has been confronted with nearly daily protests by a variety of groups, the police force has been badly weakened by mass desertions and the firing of top officials, and provincial government offices remain dysfunctional. The judicial system is hobbled by its links to the ousted regime.

The head of a Tunisian government commission on political reform warned on Feb. 22 that the country risked falling into “anarchy” as it passed through what he described as a very dangerous post-revolutionary transition toward multiparty democracy.

### **Birth of a Movement**

The Tunisian revolution began in December 2010, after a college-educated street vendor burned himself to death in protest of his dismal prospects amid Tunisia’s poverty. A wave of violent demonstrations spread, of the kind not seen since Mr. Ben Ali came to power 23 years ago in a bloodless coup. Dozens died as security forces fired on protesters.

The protesters came together after circulating calls to rally over social networks like [Facebook](#) and [Twitter](#). Many were unemployed college graduates, and they angrily demanded more jobs and denounced what they called the self-enrichment of Tunisia’s ruling family.

It is not religion, nor the adventures of a single leader, nor wars with Israel that have energized Tunisia, the subsequent uprising in Egypt and elsewhere in the region. Across

the Middle East, a somewhat nostalgic notion of a common Arab identity, intersecting with a visceral sense of what amounts to a decent life, is driving protests that have bound the region in a sense of a shared destiny.

A remarkable two-year collaboration gave birth to a new force in the Arab world—a pan-Arab youth movement dedicated to spreading democracy in a region without it. Young Egyptian and Tunisian activists brainstormed on the use of technology to evade surveillance, commiserated about torture and traded practical tips on how to stand up to rubber bullets and organize barricades.

They fused their secular expertise in social networks with a discipline culled from religious movements and combined the energy of soccer fans with the sophistication of surgeons. Breaking free from older veterans of the Arab political opposition, they relied on tactics of nonviolent resistance channeled from an American scholar through a Serbian youth brigade—but also on marketing tactics borrowed from Silicon Valley.

### **Rulers' Lavish Lifestyles Fueled Anger**

Protesters seemed to direct much of their anger at the great wealth and lavish life of President Ben Ali's second wife, Leila Trabelsi, a former hairdresser, and their extended family, most notably their son-in-law, the billionaire businessman Mohamed Sakher El Materi.

A gracious dinner at Mr. Materi's home was detailed in a [cable](#) from the American ambassador to Tunisia that was released by the anti-secrecy organization [WikiLeaks](#) and fueled at least some of the outrage: a beachfront compound decorated with Roman artifacts; ice cream and frozen yogurt flown from St. Tropez, France; a Bangladeshi butler and South African nanny; and a pet tiger in a cage.

State television reported the arrests for “crimes against Tunisia” of 33 members of Mr. Ben Ali's family. The government also said its prosecutors had opened an investigation into the family's overseas assets, while the Swiss government moved to freeze their assets in Swiss banks.

### **Dismantling a Repressive Regime**

On Feb. 22, 2011, the head of a Tunisian government commission, which is tasked with dismantling the repressive laws of the Ben Ali government, warned that the country faced a dangerous transition as it struggled toward multiparty democracy.

Yadh Ben Achour, a prominent lawyer who is the head of the country's Higher Political Reform Commission, said Tunisia would miss the two-month deadline stipulated in its Constitution for a presidential election to replace Mr. Ben Ali. It would be impossible to organize elections before the March 15 deadline, he said.

Tunisians cannot agree whether to change the current Constitution or discard it and elect a constitutional assembly that would write a new one, he said. The commission may also help draft a new constitution, a process, he said, that risked being bogged down by politicians focusing on narrow interests and not the future of the country.

#### Egypt News—Revolution and Aftermath

Egypt, the most populous country in the Arab world, erupted in mass protests in January 2011, as the revolution in [Tunisia](#) inflamed decades worth of smoldering grievances against the heavy-handed rule of President [Hosni Mubarak](#). After 18 days of angry protests and after losing of the support of the military and the United States, Mr. Mubarak resigned on Feb. 11, ending 30 years of autocratic rule. The military stepped forward and took power. It quickly suspended unpopular provisions of the constitution, even while cracking down on continuing demonstrations. On March 19, a set of constitutional amendments that pave the way for elections was overwhelmingly approved in a referendum that drew record numbers of voters. But anger over what many demonstrators saw as the military's loyalty to the core of Mr. Mubarak's government and the slow pace of change led to new mass protests and violence in April.

**April 16** [Egypt's Supreme Administrative Court officially relegated the once-supreme National Democratic Party to history](#), ruling that the party would be dissolved and its assets seized by the government.

**April 13** Egyptian authorities said the former president and his two sons [were being detained for 15 days for questioning](#) about corruption and the abuse of power during Mr. Mubarak's three-decade rule.

**April 11** Former President [Mubarak spoke for the first time](#) since being deposed in an audiotape, denying that he and his family had amassed wealth overseas and defending his honor and legacy. An Egyptian blogger was [sentenced to three years in prison](#) for criticizing the military in what human rights advocates called one of the more alarming violations of freedom of expression since the beginning of the revolution.

**April 9** Egypt's security forces [shot and killed at least two protesters and injured dozens more](#) in a predawn attempt to disperse peaceful demonstrators spending the night in the capital's iconic Tahrir Square, according to government security officials and witnesses. The crackdown was the most brutal since the overthrow of former President [Mubarak](#) on Feb. 11 and since the military started running the country.

**April 8** Tens of thousands gathered in Cairo's central Tahrir Square, waving flags and demanding the prosecution of the ousted president, [Hosni Mubarak](#), and his family in a sign of Egyptians' growing [anger with the slow pace of change under the new military rulers](#). The protest was being called the "Friday of Warning." In government ministries, factories and especially universities, [daily protests have focused on those viewed as Mr. Mubarak's surrogates](#). Demonstrators complain that the dreaded secret police vetted

every candidate for an important job under Mr. Mubarak, and that now the country deserves a clean slate.

**March 31** Egypt's military rulers announced a new interim constitution replacing the one suspended when President [Hosni Mubarak stepped down](#) on Feb. 11, incorporating the amendments approved by voters in the March 19 referendum. In addition the 18-member ruling council said it would hand over legislative powers after the parliamentary election in September 2011, and that executive powers would be transferred after the presidential election, which will be held by November.

**March 29** The military command announced that parliamentary elections would not be held until September 2011, meeting the demands of opposition leaders who wanted more time to organize political parties. The presidential election, scheduled for August, was also postponed. The command also revealed that the ousted president, [Hosni Mubarak](#), has been prohibited from leaving the country, and that it would soon lift the detested emergency law, among a number of announcements intended to shore up diminishing support for the armed forces council ruling the nation.

**March 20** Egyptian voters [overwhelmingly approved a referendum on constitutional changes](#) that will usher in rapid elections, with the results underscoring the strength of established political organizations, particularly the [Muslim Brotherhood](#), and the weakness of emerging liberal groups. More than 14.1 million voters, or 77.2 percent, approved the constitutional amendments; 4 million, or 22.8 percent, voted against them. The turnout of 41 percent among the 45 million eligible voters broke all records for recent elections, according to the Egyptian government.

## **Background**

Egypt is a heavyweight in Middle East diplomacy, in part because of its peace treaty with Israel, and as a key ally of the United States. The country, often the fulcrum on which currents in the region turn, also has one of the largest and most sophisticated security forces in the Middle East.

Mr. Mubarak has been in office since the assassination of Anwar el-Sadat on Oct. 16, 1981, whom he served as vice president. Until the recent unrest, he had firmly resisted calls to name a successor. He had also successfully negotiated complicated issues of regional security, solidified a relationship with Washington, maintained cool but correct ties with Israel and sharply suppressed Islamic fundamentalism and terrorism—along with dissent in general.

The government has maintained what it calls an Emergency Law, passed first in 1981 to combat terrorism after the assassination of Mr. Sadat. The law allows police to arrest people without charge, detain prisoners indefinitely, limit freedom of expression and assembly, and maintain a special security court.

In 2010, the government promised that it would only use the law to combat terrorism and drug trafficking, but terrorism was defined so broadly as to render that promise largely meaningless, according to human rights activists and political prisoners.

### **From Apathy to Anger**

While Mr. Mubarak's regime had become increasingly unpopular, the public long seemed mired in apathy. For years, the main opposition to his rule appeared to be the Muslim Brotherhood, which was officially banned but still commanded significant support.

In 2010, speculation rose as to whether Mr. Mubarak, who underwent gall bladder surgery that year and appeared increasingly frail, would run in the 2011 elections or seek to install his son Gamal as a successor. Mr. El Baradei, the former director of the [International Atomic Energy Agency](#), publicly challenged Mr. Mubarak's autocratic rule, but the Mubarak political machine steamrolled its way to its regular lopsided victory in a parliamentary vote.

The anger fueling the street protests was not new. It had been [seething beneath the surface](#) for many years, exploding at times, but never before in such widespread, sustained fury. The grievances are economic, social, historic and deeply personal. Egyptians often speak of their dignity, which many said has been wounded by Mr. Mubarak's monopoly on power, his iron-fisted approach to security and corruption that has been allowed to fester. Even government allies and insiders have been quick to acknowledge that the protesters have legitimate grievances that need to be addressed.

In the last few years, Egypt has struggled through a seemingly endless series of crises and setbacks. The sinking of a ferry left 1,000 mostly poor Egyptians lost at sea, an uncontrollable fire gutted the historic Parliament building, terrorists attacked Sinai resorts, labor strikes affected nearly every sector of the work force and sectarian-tinged violence erupted.

And in nearly every case, the state addressed the issue as a security matter, deploying the police, detaining suspects, dispersing crowds. That was also true in 2010, even as evidence mounted of growing tension between Egypt's Muslim majority and a Christian minority that includes about 10 percent of the approximately 80 million Egyptians.

### **A Police State**

Egypt's police bureaucracy reaches into virtually every aspect of public life here, and changing its ways is no easy task, everyone concedes. Police officers direct traffic and investigate murders, but also monitor elections and issue birth and death certificates and passports. Every day, 60,000 Egyptians visit police stations, according to the Interior Ministry. In a large, impoverished nation, the services the police provide give them wide—and, critics say, unchecked—power.

The Egyptian police have a long and notorious track record of torture and cruelty to average citizens. One case that drew widespread international condemnation involved a cell phone video of the police sodomizing a driver with a broomstick. In June 2010, Alexandria erupted in protests over the fatal beating by police of beating Khaled Said, 28. The authorities said he died choking on a clump of [marijuana](#), until a photograph emerged of his bloodied face. In December 2010, a suspect being questioned in connection with a bombing was beaten to death while in police custody. Abuse is often perpetrated by undercover plainclothes officers like the ones who confronted Mr. Said, and either ordered or allowed by their superiors, the head investigators who sit in every precinct.

The government denies there is any widespread abuse and frequently blames rogue officers for episodes of brutality. Even so, for the past 10 years, officers from the police academy have attended a human rights program organized by the [United Nations](#) and the Interior Ministry.

#### Bahrain News—The Protests (2011)

Bahrain is a small desert island kingdom in the Persian Gulf, an oil-producing nation of about 1 million that serves as a banking hub and as the base for the United State's Fifth Fleet.

In February 2011, the wave of unrest sweeping the Arab world reached Bahrain, producing a month of swelling demonstrations by the country's marginalized Shiite majority. For a time, euphoric crowds appeared on the verge of shaking serious concessions from the monarchy. Thousands filled Pearl Square in Manama, the capital, to listen to a cacophony of speeches calling for freedom.

But after a series of swings between violence and conciliation, King [Hamad bin Isa al-Khalifa](#) cracked down hard.

On March 14, the king [invited in 2,000 troops](#), including 800 from the United Arab Emirates, and 1,200, backed by tanks, from Saudi Arabia, which lies across a causeway from Bahrain. King Khalifa declared an emergency, and security forces cleared Pearl Square using bullets and tear gas. The next week, the government tore down the monument at the square's center.

Since then, Bahrain [has taken on the likeness of a police state](#). There have been mass arrests, mass firings of government workers, reports of torture and the forced resignation of the top editor of the nation's one independent newspaper.

Emergency laws give the security forces the right to search houses at will without a warrant and dissolve any organization, including legal political parties, deemed a danger to the state. Even two members of the national soccer team were arrested, despite apologizing on television for attending antigovernment rallies.

The Obama administration, which considers Bahrain a crucial ally, has issued tempered criticisms of the crackdown but has not pressed for a change in government.

In response, a once joyous but splintered opposition has been forced to come up with new strategies. The intensity of the repression is pushing some toward militancy, while others are holding back, at least for now.

The protesters were inspired by the upheavals in Tunisia and Egypt, but their struggle and challenges were different. They are predominantly Shiites, who make up 70 percent of the population, in a country whose monarchy and much of the business elite are Sunnis. While their struggle is tinged by ethnic animosity, their chief opponents in the streets are an army and security force who are predominantly foreigners, principally Pakistani, Yemeni, Iraqi and Jordanian.

The demonstrations of hundreds of thousands of people in March have given way to small marches and protests at funerals. The centers of rebellion are now in villages outside Manama like Saar and Shahrakkan, where residents have set up barricades of stones and bricks so police officers on patrol need to leave their armored cars and walk through the narrow stone pathways. Every night at 10, residents climb to their roofs and anonymously cry in protest, “God is great!”

The government and pro-government media have celebrated the relative calm in downtown Manama as a return to the kind of normality that has made this tiny island nation an important banking center and regional tourist destination. They charge that the predominantly Shiite opposition is inspired and even aided by Iran, although most Bahraini Shiites are Arabs, unlike Iranians, and associate themselves more closely with Iraqi Shiites.

The Bahraini government had originally appeared willing to compromise with the opposition, particularly the more moderate faction that wants the country to evolve into a true parliamentary monarchy that gives elected lawmakers more power. At one point last the king even apologized for the deaths of demonstrators on television, and government officials say they are still open to reform.

By the accounts of Bahraini human rights activists, 26 people have been killed, most in the past three weeks since Pearl Square was cleared. More than 300 have been imprisoned, and at least 35 people are missing.

Here is a chronology of the month of protests:

**March 18** [The Bahraini government tore down the protest movement’s defining monument](#), the pearl at the center of Pearl Square, a symbolic strike that carried a sense of finality. The official news agency described the razing as a facelift.

**March 17** The Bahraini government, which sought in February to mollify protesters clamoring for democratic reform, [decisively shifted tactics to forceful repression](#). A day after aggressively clearing Pearl Square of protesters, authorities arrested several major opposition figures, including [Hassan Mushaima](#), a Shiite and Islamist dissident politician. State television said the leaders were arrested for having “communicated with foreign countries” and because they “incited killing of citizens and destruction of public and private property.”

**March 16** Two days after the king of [Bahrain](#) brought in 2,000 troops from [Saudi Arabia](#) and other neighboring allies, and the day after he declared martial law, [his security forces rolled into Pearl Square](#), the stronghold of the antigovernment protest movement, taking it from the protesters who had moved in a month ago. Plumes of black smoke choked the central city landscape as troops repeatedly fired tear gas canisters, rubber bullets and what sounded like live ammunition, igniting fires in tents, trees and brush. Most of the hundreds in the square fled from the huge display of military might. There was no immediate word on casualties.

**March 15** Hours after the king of Bahrain declared a three-month state of emergency, doctors said [two protesters had been killed and some 200 wounded](#) and injured in clashes with riot police in the suburban village of Sitra, a stronghold of antigovernment activists six miles south of the capital. The violence contrasted starkly with a large protest in downtown Manama, where more than 10,000 protesters marched peacefully on the Saudi Arabian Embassy to denounce a military intervention by Persian Gulf countries the day before.

**March 14** [Troops crossed](#) from Saudi Arabia into Bahrain to help quell unrest there, a move Bahraini opposition groups denounced as an “occupation,” while pro-government legislators called for the imposition of martial law. The Iranian government, which has supported the protests led by the Shiite majority, branded the move “unacceptable.”

**March 13** Thousands of antigovernment protesters in Bahrain [blocked access to the financial district in Manama](#), the capital, preventing workers from getting to their offices and pushing back police officers who tried to disperse them. It was the most serious challenge to the royal family that rules Bahrain since protests began last month. Witnesses said the police used tear gas and fired on the protesters with rubber bullets.

**March 12** Defense Secretary [Robert M. Gates](#) warned this tiny kingdom’s ruling family that [“baby steps” toward reform would not be enough](#) to meet the political and economic grievances sweeping the region. Mr. Gates also cautioned Bahrain’s king and crown prince during two hours of meetings in Manama, Bahrain’s capital, that if the reform process was prolonged, the United States feared that Iran would become involved and create more chaos.

**March 6** The leaders of Bahrain’s opposition movement said that they [would not be mollified by offers](#) of money and jobs, raising the prospect of a protracted standoff

between protesters and the embattled government of this strategically important Persian Gulf island nation.

**March 4** Thousands of Shiite protesters [converged on the state television headquarters](#) outside the capital, Manama, The Associated Press reported. The latest protest came a day after brief clashes between groups of Sunnis and Shiites in the town of Hamad that again highlighted the sectarian divisions that have been driving demonstrations in the tiny Persian Gulf nation.

**Feb. 27** The leader of a banned opposition party, Hassan Mushaima, [returned from exile and exhorted a crowd of tens of thousands](#) of antigovernment protesters to continue demonstrating until they achieved a “successful revolution.” Mr. Mushaima, a dissident who has long demanded fundamental changes in what is effectively an absolute monarchy, appeared to be trying to move into a leadership role in the opposition movement.

**Feb. 25** Pro-democracy [demonstrations on a scale that appeared to dwarf the largest ever seen](#) in the tiny Persian Gulf nation blocked miles of downtown roads and highways in Manama, overflowing from Pearl Square for the second time in a week. For the first time however, it was the country’s Shiite religious leaders, rather than the political opposition, who called for people to take to the streets. Although some of the chants and symbols had a religious cast, protesters’ demands remained the same, emphasizing a nonsectarian call for democracy and the downfall of the government.

**Feb. 22** [More than a hundred thousand protesters poured into Pearl Square](#) in an unbroken stream stretching back for miles along a central highway in the biggest antigovernment demonstration yet seen in this tiny Persian Gulf kingdom. Security forces were nowhere to be seen along the demonstration route. The protesters streaming into the square on Tuesday joined thousands of others who have camped out in order to occupy the area after the military pulled out following a deadly crackdown last week. The sheer size of the gathering was astonishing, and the new arrivals were likely to overflow into the area surrounding the square.

**Feb. 21** The [organizers of the Bahrain Grand Prix announced](#) that the Formula One race which was scheduled to run March 13 at the Sakhir circuit, outside the capital, Manama, had been canceled because of the political unrest in the country. Bahrain has held the race since 2004 at a purpose-built circuit about 50 kilometers, or 30 miles, from the capital city. But the majority of the teams, media and foreign spectators stay in hotels in the center of the city, many of them near Pearl Square, which has been the central meeting point of the demonstrations.

**Feb. 20** Teachers, lawyers and engineers again marched into Pearl Square, joining an emboldened opposition whose [political leaders demanded](#) that King [Hamad Bin Isa al-Khalifa](#) dissolve the government and fire his uncle, who has held the post of prime minister for 40 years, before they agree to enter into talks. The crown prince, Salman bin

Hamad al-Khalifa, continued to call for a national dialogue and period of mourning as a step toward reconciliation. But officials with the leading opposition party said that with six people dead, hundreds injured and many still missing, the king would first have to convince them that he was willing to compromise after repeatedly unleashing his security forces on civilians.

**Feb. 19** Thousands of [jubilant protesters surged back into Pearl Square](#), the symbolic heart of Bahrain, after the government withdrew its security forces, calling for calm after days of violent crackdowns. The shift was at least a temporary victory for the Shiite protesters, who had rejected a call to negotiate from Bahrain's Sunni monarch until the authorities pulled the military off the streets. A review of administration statements shows that [American officials overlooked recent complaints](#) about human rights abuses in the strategically important kingdom.

**Feb. 18** [Government forces opened fire on hundreds of mourners](#) marching toward Pearl Square. A Western official quoted a witness as saying that the shooters were from the military, not the police, indicating a hardening of the government's stance against those trying to stage a popular revolt. [Video: Bahrain: A Diplomatic Puzzle](#)

**Feb. 17** [Hundreds of heavily armed riot police officers rushed into Pearl Square](#), firing shotguns, tear gas and concussion grenades at the thousands of demonstrators who were sleeping there. At least five people died, and at least 200 were wounded, according to medical personnel. At the main hospital following the violence, thousands gathered screaming, crying and collapsing in grief.

**Feb. 16** [Thousands of protesters poured into this nation's symbolic center](#), Pearl Square, in a raucous rally mirroring the unrest convulsing, if not transforming, governments of the Middle East. The crowd grew bolder as it grew larger, and modest concessions from the government seemed only to raise expectations among the protesters.

## **A Troubled History**

Though the al-Khalifa family has ruled since the 18th century, Bahrain has been a politically troubled nation for generations.

The kingdom gained independence from Britain in 1971. Soon after, the price of oil spiked during the Arab-Israeli war of 1973. Bahrain's oil reserves were small compared to its neighbors', but provided fuel for rapid development.

For a period, the island kingdom was hailed as a model of reform in the Arab world. King Hamad ended a 25-year state of emergency in 1999 and promised to turn Bahrain into a constitutional monarchy with a newly elected Parliament.

But in 2001, he reneged on his promise and imposed a constitution establishing a second, appointed legislative house that significantly curtailed the power of the elected one.

Gerrymandering ensured a Sunni advantage in the elected house, too, prompting the predominantly Shiite opposition to boycott elections in 2002.

In the years since, Bahrain's Shiites have been marginalized. They allege that the government has backed Sunni Islamists and encouraged increasingly sectarian politics. Activists in the country say they have uncovered a systematic effort to naturalize Sunnis from other Arab countries as well as from Southeast Asia.

Shiites are all but banned from the military and security forces—certainly from command positions—one of their primary grievances. The police force is staffed primarily by foreigners: Syrians, Iraqis, Jordanians, almost anyone who happens to be a Sunni and is eager to earn a Bahraini passport.

## **Background**

There has long been tension between the Sunni Muslim king, [Hamad Bin Isa al-Khalifa](#), the royal family and ruling elites, and the approximately 70 percent of the local population that is Shiite. About half the residents of Bahrain are foreign workers.

Since late 2009, Shiites in villages around the nation have been holding regular protests, burning tires in the road and demanding the release of dozens of political prisoners.

In February 2011, [skirmishes broke out between young protesters and heavily armed government forces](#) in villages around Manama, the capital. The clashes appeared to have rattled the leadership into trying both enticement and fear. King Hamad announced that the state was giving every Bahraini family the equivalent of \$2,700 in cash, and he filled the streets with heavily armed riot police.

Tens of thousands of demonstrators filled Pearl Square in Manama, camping out in an almost carnival-like atmosphere and calling on the king for reforms.

The pro-democracy protests in this island nation have followed the pattern of those in Egypt and Tunisia, with cell phones and [Facebook](#) posts propelling the movement and a botched, deadly crackdown by security forces serving to embolden the demonstrators.

Yet those who lead and take part in the nearly daily demonstrations say they [fear at least one key difference](#): The United States may not be fully on their side.

Unlike in the case of Egypt, where [President Obama](#) promised to “stand up for democracy” and called for a change of power ~~now~~,” Washington has backed the royal family in Bahrain with statements supporting the country's still-undefined proposal for dialogue with the opposition.

Obama administration officials say they believe the royal family has earned the right to try to navigate this period, after heeding the United States' plea to call off the security forces who shot the protesters, killing seven of them.

## Jordan

Jordan, one of America's most important allies in the Middle East, was hit in late January 2011 by the waves of unrest that have spread across the Arab world in the wake of the revolution in [Tunisia](#). On Feb. 1, King [Abdullah II](#) dismissed his cabinet and prime minister in a surprise move meant to calm street protests that have also been fueled by the country's worst economic crisis in years.

The royal palace announced that the king had dismissed Prime Minister Samir Rifai and replaced him with Marouf al-Bakhit, who has served before in the post and is a former general and a onetime ambassador to Israel and Turkey widely viewed as clean of corruption.

The rest of the new government was sworn in on Feb. 9. The 27-member cabinet included a number of holdovers—notably the foreign, interior and economy ministers—as well as a few leftists. The opposition Muslim Brotherhood, which declined an invitation to be part of the government, said it was less interested in who held the portfolios than what they would do.

Demonstration and counter-demonstrations followed but remained peaceful until March 24, when at least one man died when government supporters attacked a tent camp that pro-democracy protesters had set up in the center of Amman, the Jordanian capital, in conscious imitation of Tahrir Square in Cairo. The violence stopped when security forces intervened, and a week later a new round of demonstrations went off peacefully.

Changing cabinets is not new for King Abdullah. In his 12 years on the throne, he has done so eight times. But this was the first time that he had done so in reaction to public pressure, seeking to undermine a growing protest movement across a broad spectrum of society and to pre-empt further unrest. It came after four weeks of unusual public demonstrations.

Buffeted by the forces at play across the region—rising prices, a bulging underemployed youth population, the rapid spread of information and resentment, an unaccountable autocracy—Jordan is on edge. The promises made by the king when he fired his cabinet seem to have bought him some time. But many question whether the promises will be fulfilled, and whether such steps will in any case be enough to calm the rising tide of frustration.

What is most striking right now in Jordan is that the very system of the monarchy seems open to question. This is partly because of what is happening elsewhere in the region but also because of growing discontent with King Abdullah and his wife, Queen Rania. King

Abdullah's father, King Hussein, who ruled for 46 years, enjoyed near adoration of his people.

The concerns and complaints of the different constituencies in Yemen are not only distinct, they are often contradictory, and the monarchy faces little risk of opponents coalescing into the kind of mass movements seen in Tunisia and Egypt.

Jordan's main constituencies are the so-called East Bankers or tribes, and the [Palestinians](#) who constitute a majority of the nation's six million people. East Bankers, the country's original inhabitants, dominate the civil service, especially the security forces, while the Palestinians rule in the private sector. Economic reform to bring Jordan in line with the global marketplace has tended to benefit the Palestinians, while the East Bankers—the core of the monarchy's support—rely on the government payroll.

In addition, the king maintains his distance from the complaints by allowing blame to fall on government ministers, whom he replaces at will.

Jordan is a highly literate and largely stable country, with well-developed security and intelligence operations. But it has a fundamental vulnerability in the large number of Palestinians living there. Refugees arrived in large numbers from the West Bank and Jerusalem after the war in 1967, and more arrived from Kuwait after President [Saddam Hussein](#) of Iraq invaded that country in 1990. They and their descendants make up nearly half the country's population of six million.

The recent demonstrations in Jordan were the first serious challenge to the rule of King Abdullah, a crucial American ally who is contending with his country's worst economic crisis in years.

On Jan. 28, thousands took to the streets in the capital, Amman, as well as several other cities shouting, "We want change!" Because direct criticism of the king is banned, the focus has been on his government. Banners decried high food and fuel prices and demanded the resignation of the prime minister, appointed by the king.

Journalists, students and retired generals have taken the unusual step in recent months of attacking corruption, restrictions on freedom of political expression and reductions in government subsidies.

The mass protests on Jan. 28 were led by the Islamic Action Front, but included leftists and trade unions. Earlier that week, thousands marched in a similar protest, demanding the right to elect the prime minister, who is currently appointed by the king.

The demonstrators also protested economic hardship, a common refrain across the region. In a bid to tamp down some of the frustration, the king announced \$125 million in subsidies for basic goods and fuel and an increase in civil servant pay.

In November 2010, parliamentary elections were held but were boycotted by the Islamic Action Front, which said the king had gerrymandered the parliamentary districts to increase the representation of the rural areas where he finds his strongest support.

The campaign was dominated by widespread anger at the stalled Middle East peace talks and concerns about the rising poverty rate. The pro-American kingdom has struggled for years to balance a measure of democracy with the need to check a powerful Islamist movement and a large, restive [Palestinian](#) population.

Abdullah's ascension to the throne in 1999 came as a surprise: his father named Abdullah as his successor only two weeks before his death. The next king was supposed to be Crown Prince Hassan, youngest brother of King Hussein. Prince Abdullah had a reputation as a bit of a lightweight, a Prince Hal with a vaguely Falstaffian cast of friends who drove fast cars and enjoyed the company of women. But a year after ascending the throne, the king had reached a level of popularity no one in Jordan could have reasonably expected.

#### **Oil above \$112 amid Mideast turmoil, weak dollar**

(AP) SINGAPORE (AP)—Oil prices fell slightly to above \$112 a barrel Monday in Asia as traders eyed fresh Middle East tension and a wobbly U.S. dollar.

Benchmark crude for May delivery was down 42 cents at \$112.37 a barrel at late afternoon Singapore time in electronic trading on the New York Mercantile Exchange. On Friday, the contract jumped \$2.49 to \$112.79, the highest since September 2008.

In London, Brent crude for May delivery was down 62 cents to \$126.03 a barrel on the ICE Futures exchange.

Oil prices have soared about 33 percent since mid-February as traders worry political violence in the Middle East and North Africa could disrupt crude supplies.

Violence escalated last week between Palestinians in Gaza and Israel. Since Thursday, Palestinians have fired more than 120 rockets and mortar shells into southern Israel, prompting Israeli reprisals that have killed 19 Palestinians, the most intense fighting between Israel and Gaza militants since January 2009.

Meanwhile, Egyptian soldiers Saturday attacked protesters calling for an investigation of former President Hosni Mubarak for embezzlement, killing at least one person and injuring 71 others. Several hundred protesters remained barricaded at Cairo's Tahrir Square.

—Fresh headlines over the weekend could portend another difficult week for oil bears,” energy consultant The Schork Group said. —Gaza-Israel violence along with new protests

in Tahrir Square against the military could incite another buying frenzy in the market.”

Investors are also watching closely the currency markets as the U.S. dollar fell to a 15-month low against the euro last week. A weaker U.S. currency makes dollar-based commodities such as oil cheaper for investors with other currencies.

The euro was down slightly at \$1.4457 on Monday from \$1.4483 late Friday.

Analysts expect the surge in oil prices will undermine consumer demand, but some are optimistic higher fuel costs won't derail the global economy recovery.

—The rise in the price of oil in reaction to the spreading of unrest from Tunisia and Egypt to Libya and Bahrain remains below the threshold that is likely to have lasting impact on the global economy,” said Jeffrey Morrison of MFS Investment Management. —We expect the global economic recovery to be sustained.”

In other Nymex trading in May contracts, heating oil fell 1.8 cents to \$3.30 a gallon and gasoline dropped 1.4 cents to \$3.25 a gallon. Natural gas futures were down 2.6 cents at \$4.02 per 1,000 cubic feet.

### **Oil Prices and Middle East Turmoil: The Economic Consequences**

[Christopher Boucek](#), Jamie Webster, [Hans Timmer](#), [Uri Dadush](#), Mohsin Khan Thursday, March 31, 2011 – Washington, D.C.

#### **Resources**

Even though recent unrest in the Middle East has been limited to countries that have little effect on the global supply of oil, uncertainty in the oil markets has surged and prices have increased significantly. If prices rise further and stay elevated for a period of time, the global recovery could be in danger.

Carnegie's Christopher Boucek, Mohsin Khan of the Peterson Institute, the World Bank's Hans Timmer, and PFC Energy's Jamie Webster discussed the complex set of political and economic issues shaping today's oil markets. Carnegie's Uri Dadush moderated.

#### **Today's Oil Price**

Even before turmoil broke out in the Middle East, the global recovery—and rising demand from emerging markets in particular—had placed upward pressure on oil prices. The Brent Crude oil marker, which is used to price the majority of crude oil, was at \$90 per barrel in December. Since then, little has changed in the oil market's supply and demand, but, Webster explained, market fears have caused Brent to surge to \$115 per barrel.

- **Market Fundamentals:** Webster explained that the biggest changes in the Middle East have so far occurred in countries that have only modest effect on the actual oil market. Tunisia is neither a large supplier nor consumer of oil; Egypt's Suez Canal and Sumed Pipeline have not posed a transit risk; and the disruption in

Libya's production poses only a modest production risk. Furthermore, Saudi Arabia and the Emirates, which have not seen widespread protests, have unused oil capacity that can offset the decline in Libyan exports. Based on market fundamentals, Webster said, the oil price should be closer to \$80 or \$90 per barrel—near the lowest price Saudi Arabia requires to cover all its government spending.

- **Market Fears:** Oil prices are significantly higher, however, because investors are worried that unrest could erupt in Saudi Arabia, the region's biggest oil producer and supplier of reserves. According to Boucek, however, revolution is not likely there. Saudi Arabia is deploying a powerful mix of force, religious ideology, and public spending to quell unrest.

Though the oil price has risen, Timmer noted that price volatility was low in March. Even as events unfolded in the Middle East, the price held steady, suggesting that forward-looking oil markets had already taken the unrest into account. Feedback mechanisms that keep the price from surging too high and a healthy oil supply may also have helped.

### **Short-Term Concerns**

If another disruption affects supply, prices could spike very quickly, with potentially dangerous implications, the panelists warned. Khan cited research by the International Monetary Fund, which determined that a sustained price of \$120 per barrel—not far from the Brent marker's recent \$115 per barrel—would inflict some damage on the world economy. Under a plausible bad-case scenario—disruptions in Algeria's oil supply, for example—the price could even reach \$140 or \$150 per barrel over an extended period, enough to cut a percent or more from world growth. The panelists agreed that estimates are more likely to undershoot than overshoot the realized oil price.

- **Political Worries:** From a political standpoint, Yemen and Bahrain are the biggest concerns and could spell trouble for oil markets—the first through transport obstructions and the second through its proximity to Saudi Arabia, explained Boucek. Regional tensions could ignite fears that Iran blocks the Strait of Hormuz, though, as Webster noted, Iran depends vitally on the Strait for transporting oil. Boucek added that some of the region's players, such as al-Qaeda, have a history of attacking oil infrastructure.
- **Macroeconomic Implications:** Timmer noted that any increase in the oil price will be harmful. For every 50 percent increase in the oil price, global economic growth will diminish by 1.5 percent. This will prove particularly difficult for high-income countries, which are already growing at lower rates than emerging

economies, and especially for Euro area countries already facing debt crises. Headline inflation could also spiral and encourage premature monetary tightening, particularly in developing countries, which could have secondary effects on financial markets. Most dangerously, a higher oil price could hurt production capacity and lead to another recession, Timmer warned.

- **Poverty and Employment:** In addition, the higher price of oil and their spillover onto food could push millions of people into deeper poverty, said Timmer. Khan predicted that employment in high-income countries would also suffer significantly.

### Looking Ahead

Panelists agreed that demand for oil is likely to rise rapidly and perhaps outpace supply over the next three to four years even without accounting for supply disruption in the Middle East.

- **Demand from Asia:** Though China's most recent Five Year Plan aims to lower its oil consumption, Khan argued that both China and India will grow as oil markets as they build up their strategic oil reserves. Meanwhile, the damage done to Japan's Fukushima nuclear plants by the recent earthquake and subsequent tsunami could turn the entire discussion of nuclear power on its head, reducing the likelihood that nuclear energy's role as an alternative to oil increases. Timmer suggested that general energy and climate policy will affect oil prices.
- **The Middle East:** Khan suggested that developments in the Middle East will be particularly interesting to watch. He predicted a rise in populism in the region, with expansionary government policy to provide more jobs and food and fuel subsidies to citizens. Timmer agreed that the changes will be very important, as stability will emerge only when unemployment in the region is addressed. Webster suggested that Saudi Arabia, which uses crude oil for power, could become an even larger oil consumer in the future.
- **Supply:** Webster noted that Iraq's oil production will likely increase, particularly in the short run, though he predicted protests could escalate in August and potentially offset the oil supply growth. Khan added that there is a need to set up more refineries capable of processing heavy Saudi oil to ensure that its unused supply capacity is effective in moderating incipient price pressures.

### G20 Policy

There is very little that the major economies can do to affect the oil market in the short term, but they can change the game in the long term by lowering their reliance on fossil fuels and instituting carbon taxes.

- **Strategic Oil Reserves:** The release of oil reserves to lower prices is unlikely, stated Khan. Most countries are actually moving in the opposite direction and building reserves. In addition, Webster argued that releasing reserves would likely have little impact—prices are not reflecting a lack of oil supply but rather a host of political and economic concerns.
- **What Not to Do:** Timmer noted that countries should refrain from decreasing gas taxes, subsidizing energy use to protect their own economies, or portraying climate change policy as only a long-term solution. Because oil markets are very forward-looking, announcing bold, realistic policies to mitigate climate change could dampen oil price pressures in the short term.

### **Afternoon Price Check: Oil Prices Rally On Libya Fighting, Middle East Protests**

Posted by Josh Garrett on March 22, 2011 at 4:10 pm

Civil war in Libya and unrest and uncertainty in the Middle East drove crude oil prices higher at the NYMEX today. (image: ft.com)

World oil markets now appear to have re-focused completely on the events in the Middle East and North Africa and their implications for global crude supplies. International intervention in Libya could prolong the civil war there, which would mean an extended loss of at least a million barrels per day of crude oil supplies to the world. As street protests and violent crackdowns on activists in Bahrain, Yemen, and Syria continue, nervousness on oil markets continues to build. The primary concern is that the political movements could spread to Saudi Arabia, the world's leading oil exporter, and take a substantial bite out of oil supplies. Turbulent times in Libya in the Middle East drove oil prices higher at the NYMEX today, overshadowing an announcement that Japan will tap its strategic crude oil reserve to help power recovery efforts. That announcement, which implied a major (but temporary) reduction in global demand, pushed prices lower in early trading before giving way to gains by the afternoon. Crude oil closed at its highest price since the March 11 earthquake in Japan. Today's price increases at the NYMEX will likely bring a small to moderate increase in retail heating oil prices tomorrow.

### **What impact will the Middle Eastern protests have on the world economy?**

By

[James Arter](#)

– March 31, 2011

The Middle Eastern protests have dominated the news agenda ever since the ‘Jasmine revolution’ began in Tunisia in mid December. The protests have since spread to many neighbouring countries; there have been protests as far afield as Mauritania, Somalia and Iran. Libya is in the throes of a conflict approaching civil war as protesters opposed to Colonel Gaddafi have seized key cities such as Benghazi and Tobruk and are attempting to gain control of the capital Tripoli. The protesters have undoubtedly caused a seismic shift in the political landscape of the Middle East.

Like most global events of this significance, it is already becoming apparent that the turmoil in the area is affecting the global economy. The Middle East is a major source of global oil production with Libya, Bahrain, Yemen, Iran and Algeria together supplying a tenth of the world’s oil. Additionally Saudi Arabia, as the regional heavyweight, accounts for 8.5 million barrels per day, making it the world’s third biggest supplier.

Following the unrest in many of these nations, the global price of oil, which had seen significant increases in the months before the protests, has rocketed to a high of \$119 per barrel, the highest for two and a half years and the price seems set to rise with further unrest predicted. The market worries that oil production in the region could shrink due to the workforce either protesting or having fled to avoid the violence. Additionally, protests increase the price of future extraction. Investors would add a risk premium to any new projects in the area, making investment more expensive. Furthermore, markets don’t like sudden changes; the speedy ousting of governments across the region and widespread protests appear to have spooked the market, raising fears of another oil shock similar to the first gulf war or the Iranian revolution.

Increasing oil prices directly affect consumers through higher fuel prices at the pump as well as additional costs through industries that use large quantities of oil as inputs or fuel. For example, last week both Thompson and Thomas Cook both added additional fuel surcharges onto the costs of flights in response to the increased price of fuel. Any increase in oil prices will have a knock on effect on general transport costs, for instance the cost of shipping food around the globe is likely to increase leading to further rises in global food prices. Food prices are currently already rising as a result of last year’s droughts in Russia and China and heavy floods in Pakistan.

The current inflation rate in the UK is 4%, fully twice the target rate; any increase in fuel and food prices is likely to push this even higher. This will have significant ramifications for the wider economy. A higher inflation rate reduces the real wages of workers lowering overall spending within the economy. With the UK teetering on the edge of a ‘double-dip’ recession any decrease in consumer spending may weaken the struggling economic recovery.

However currently, disruption to oil supply is relatively slight. Libya for instance only produces about 2% of the world’s oil and other countries such as Algeria and Bahrain supply even less. If Saudi Arabia were to be affected by similar protests as have been seen in the rest of the region then it is likely that the effects would be much more serious. The Saudi leaders have promised reform in an attempt to placate the protesters and applied pressure on its neighbour Bahrain to take a strong line with its own protesters lest the protests spill over the border into the country. However a ‘day of rage’ has been

called for the 11<sup>th</sup> of March in the capital Riyadh and the markets are understandably nervous as to the extent of the protest and any disruption caused to oil output.

However the effects of the protests have been felt most keenly by the Middle Eastern countries themselves. With millions of people off work, foreign companies evacuating staff and the slowdown in commerce, the economies of the region almost halted. However long after the short term damage is repaired, the region's international reputation will be tarnished. Rioting and large scale civil unrest do not attract business and unity governments or those thrust into power on the back of popular protest may be weak and leave the area prone to slip back into chaos. In recent years countries like Bahrain and the UAE have attracted huge levels of foreign investment with cities such as Dubai booming with vast sums of money. The recent violent clashes in Bahrain, where at least seven people have died, have dashed the image as the area being a 'business idyll'. With the prospect of further conflicts in the area, it is likely that the Gulf States will find it harder to attract foreign companies and investment in the future. This could be devastating to the smaller states' economies that rely on business tax revenues rather than the larger states with significant oil reserves.

Tourism is a key part of many of the regions' economies, for example 5.5% of Tunisia's income comes from tourism and the sector employs one in five of the workforce. The violence has forced many tourists to leave the area and at least in the short term it is unlikely that tourists will return until the area's reputation improves.

The global stock markets have also not been spared the effects of the instability in the area, with many stock markets falling as the severity of the unrest increased and uncertainty as to the length and impact of the protests grew. Companies that are exposed to oil price fluctuations have suffered the greatest with many airlines, for example, suffering falls in their share price as their costs rose.

Undoubtedly the recent events in the Middle East offer great hope for the people of the region but will also be highly influential for the global economy. At present it seems like there will be a moderate increase in oil prices and the effects of such an increase will be amplified through higher prices for a variety of fuel dependent industries. These will have an effect on consumer spending and inflation and so have a knock-on effect on each country's macroeconomic performance. Additionally the protests will have been felt across the business world with volatility in the stock market and lower investment in Middle Eastern countries. However the long term effects remain to be seen and depend heavily on what happens in the next few months. Were Saudi Arabia or some other major oil exporter like Iran to suffer some of the turmoil that their neighbours have, then the global effect of the protests could be far larger than initially predicted. With Libya still in a state of violent flux it is impossible to say what may happen into the future. If the violence was to continue for any length of time or result in any serious damage to the oil infrastructure in the country such as the export terminals in Tobruk, Brega or Tripoli then the world oil price would rise to new heights.

### **Libya crisis drives up oil prices**

Clifford Krauss & Jad Mouawad, New York Times, Mar 2, 2011, 01.45pm IST

HOUSTON: Just when oil markets appeared to be calming, crude oil prices surged again on Tuesday as the potential for more oil shipment disruptions spread across the [Middle East](#) and [North Africa](#).

With Libya's oil exports almost entirely halted for the last several days, renewed unrest in Oman, [Iran](#) and [Iraq](#) rattled oil traders. An interruption of shipments from any of those countries would further tighten oil supplies, even as [Saudi Arabia](#) has rushed to fill the vacuum of Libyan supplies by pumping more oil from its fields.

The worries about the oil supply rippled through other markets, with stock markets turning lower on concerns that the higher cost of energy would slow economic recovery. Gold prices also surged on the latest reports, and indexes on Wall Street declined sharply, with the Dow Jones industrial average down more than 1.3 percent. The Saudi Arabian benchmark stock index fell 6.8 percent.

In the latest sign that the political contagion was spreading, demonstrators in [Oman](#) on Tuesday tried to block a major road leading to the industrial port town of Sohar. Protesters in recent days have set fire to at least one police station and two government office buildings in the normally stable Persian Gulf country, which is ruled by a family dynasty and is the largest non-OPEC oil producer in the Middle East.

—To have protests in Oman, which had previously been seen as a sleepy gulf kingdom, heightens concerns that nowhere is immune from the contagion affects,” said Helima L. Croft, a director and senior geopolitical analyst at Barclays Capital. —Every day we seem to have a new country with a new problem.”

Oman produces 860,000 barrels of oil daily, almost 1 percent of world supplies, and its production has been rising in recent years with investments from [Royal Dutch Shell](#), BP, Repsol and other international companies. Its importance is magnified by the fact that its crude is of such quality that it can be blended by most refineries around the world, although most of its exports now go to [China](#) and Japan.

Oman straddles the [Strait of Hormuz](#), a strategic route through which 40 percent of the world's oil tanker traffic crosses. On the other side of the strait lies Iran, another major producer, where there were reports on Tuesday that security forces had used tear gas to disperse protesters in [Tehran](#). Iran, with approximately 10 percent of the world's oil reserves, exports about 3.7 million barrels a day

### **Middle East and North African Turmoil Raised the Oil Prices**

Fluctuating dollar and hostile environment in the Middle East including political instability in North Africa added to the worries of the [traders](#) taking oil prices up by jump of 33% since [February to \\$112 a barrel on Monday](#).

During the late afternoon [trading](#) on the New York [Mercantile Exchange](#), the prices of the crude for the May delivery slipped by 42 cents from \$112.79, highest since September 2008, to \$112.37 per barrel.

In middle east Asia, firing of 120 rockets and mortar shells by Palestine into southern Israel has once again intensified the violence between the two countries while in northern Africa, the Egyptian soldiers assaulted the protestors and killed one, who were demanding the investigation for the embezzlement of funds by former President Hosni Mubarak.

On Monday, the euro slipped slightly to \$1.4457 from \$1.4483 of late Friday. Jeffrey Morrison of MFS Investment Management asserted, “The rise in the price of oil in reaction to the spreading of unrest from Tunisia and Egypt to Libya and Bahrain remains below the threshold that is likely to have lasting impact on the global economy.” The U.S. dollar fell to a 15-month low against the euro, thus making the dollar based commodities cheaper for the euro based investors.

THIS PAGE INTENTIONALLY LEFT BLANK

## APPENDIX D. EXAMPLE XML FILE GENERATED BY SPIDER

```

<?xml version="1.0"?>
-<gov.sandia.cortext.document.Document>    <string>7917ffe9-608e-49e5-9b35-2cccd1dfa63a</string>
<int>2</int>    <string>content</string>    <string>enctype=application/x-www-form-urlencoded"
id=doc_search_form"> <!-- name not valid in XHTML, To get around this issue, use <form
id=header_search_form"    ...    >    and    in    my    link    use
javascript:document.forms[_header_search_form'].submit(); which validates correctly. --> <fieldset
class=rightAligned"> <p><input class=query csx" type=text" name=q" value= size=50" /> <input
class=button csx" type=submit" name=submit" value=Search" alt=Search" /></p> <input
type=hidden" name=sort" value=fel" /> <p><input type=checkbox" name=ie" value=1" /> Include
Citations &#124; <a href=/advanced_search.jsessionid=5AC149882DBB363288302750A37F33E6"
title=Search full text, title, abstract, date, author name, author affiliation, etc.">Advanced Search</a>
&#124; <a href=/help/search.jsessionid=5AC149882DBB363288302750A37F33E6" title=Get
Help">Help</a></p> </fieldset> </form> </div> <!-- End docs_tab --> <div id=auth_tab"
class=mootabs_panel">    <form    method=get"
action=/search.jsessionid=5AC149882DBB363288302750A37F33E6"    enctype=application/x-www-
form-urlencoded" id=auth_search_form"> <!-- name not valid in XHTML, To get around this issue, use
<form    id=header_search_form"    ...    >    and    in    my    link    use
javascript:document.forms[_header_search_form'].submit(); which validates correctly. --> <fieldset
class=rightAligned"> <p><input class=query csx" type=text" name=q" value= size=50" /> <input
class=button csx" type=submit" name=submit" value=Search" alt=Search" /> <input type=hidden"
name=sort" value=eite" /> <input type=hidden" name=t" value=auth" /></p> <p> <input
type=checkbox" name=auth" value=1" checked/> Disambiguated Search &#124; <input
type=checkbox" name=ie" value=1" /> Include Citations &#124; <a
href=/advanced_search.jsessionid=5AC149882DBB363288302750A37F33E6" title=Search full text,
title, abstract, date, author name, author affiliation, etc.">Advanced Search</a> &#124; <a
href=/help/search.jsessionid=5AC149882DBB363288302750A37F33E6" title=Get Help">Help</a></p>
</fieldset> </form> </div> <!-- End auth_tab --> <div id=table_tab" class=mootabs_panel"> <form
method=get"    action=/search.jsessionid=5AC149882DBB363288302750A37F33E6"
enctype=application/x-www-form-urlencoded" id=table_search_form"> <!-- name not valid in XHTML,
To get around this issue, use <form id=header_search_form" ... > and in my link use
javascript:document.forms[_header_search_form'].submit(); which validates correctly. --> <fieldset
class=rightAligned"> <p><input class=query csx" type=text" name=q" value= size=50" /> <input
class=button csx" type=submit" name=submit" value=Search" alt=Search" /> <input type=hidden"
name=t"    value=table"    /></p> <p><a
href=/help/search.jsessionid=5AC149882DBB363288302750A37F33E6" title=Get Help">Help</a></p>
</fieldset> </form> </div> <!-- End table_tab --> </div> <!-- End search_tabs --> <!-- End
search_meta_box --> <script type=text/javascript"> <!-- window.addEventListener(_domready', function(){
startSearchTabs(_docs_tab'); }); // --> </script> <!-- End of the floating containers --> <!-- End header for
the primary, secondary, tertiary, etc. pages --> <div id=page_wrapper" class=clearfix"> <!-- Contains all
the divisions (div's) within the page (not including top navigation bar, search box and meta information) --
> <div id=eenter_content" class=clearfix"> <!-- Contains header div --> <div
id=secondary_tabs_container"> <div id=secondary_tabs"> <!-- to make a tab highlighted, place
id=eurrentpage" within li tag --> <ul><li id=eurrentpage"><a class=page_tabs remove"
href=/viewdoc/summary.jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750"><s
pan>Summary</span></a></li><li    class=tab4sublist"><a    class=page_tabs    remove"
href=/viewdoc/similar.jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750"
title=Find documents related to this document"><span class=clearfix"><span class=tab_sublist    clearfix"><li><a class=sublist_links remove"
href=/viewdoc/similar.jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750&amp;
type=ab" title=Documents that cite the same works"><span>Active Bibliography</span></a></li><li><a
class=sublist_links

```



mp;title=Memory Performance of Master Go Players&site=citeulike" title="CiteULike">D.: Perception and memory versus thought: Some old ideas and recent findings</a> &ndash; Groot - 1966 </td></tr><tr><td> <span class=ehar\_emphasized">11</span> </td> <td> <a class=eatation\_only" href=/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=1320987">The impact of chess research on cognitive science</a> &ndash; Charness - 1992 </td></tr><tr><td> <span class=ehar\_emphasized">8</span> </td> <td> <a class=eatation\_only" href=/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=1617061">An analysis of strong Go-players&#039; protocols</a> &ndash; Saito, Yoshikawa - 1996 </td></tr><tr><td> <span class=ehar\_emphasized">6</span> </td> <td> <a class=eatation\_only" href=/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=3455954">Aspects of Human Problem Solving: The Use of Internal Representations</a> &ndash; Eisenstadt, Kareev - 1975 </td></tr><tr><td> <span class=ehar\_emphasized">5</span> </td> <td> <a class=eatation\_only" href=/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=1617060">Do Go players think in words? interim report of the analysis of Go players&#039; protocol</a> &ndash; Saito, Yoshikawa - 1995 </td></tr><tr><td> <span class=ehar\_emphasized">4</span> </td> <td> <a class=eatation\_only" href=/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=4964019">Diifferent ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks</a> &ndash; Humphreys, Bain - 1989 </td></tr><tr><td> <span class=ehar\_emphasized">2</span> </td> <td> <a href=/viewdoc/summary;jsessionid=5AC149882DBB363288302750A37F33E6?cid=1957941">The Use of Inferential Information in Remembering Go Positions</a> &ndash; Burmeister, Wiles - 1996 </td></tr><tr><td> <span class=ehar\_emphasized">2</span> </td> <td> <a class=eatation\_only" href=/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=5014568">Groot: Thought and choice</a> &ndash; de - 1965 </td></tr><tr><td> <span class=ehar\_emphasized">1</span> </td> <td> <a class=eatation\_only" href=/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=7209545">Deducing memory structures from inter-response times</a> &ndash; Reitman - 1976 </td></tr><tr><td> <span class=ehar\_emphasized">1</span> </td> <td> <a class=eatation\_only" href=/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=7209546">Go as a testbed for cognitive science studies. To appear in IJCAI workshop proceedings, Using Games as an Experimental Testbed for AI Research</a> &ndash; Saito, Yoshikawa - 1997 </td></tr></table> <div class=parent\_div\_spacer"></div> </div> <!-- End main\_content --> <div class=information\_bar"> <div class=para4"><a href=/viewdoc/download;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750&amp;rep=rep1&amp;type=pdf" title=View or Download this Document">View or Download</a> &#124; <a href=/myciteseer/action/addPaperCollection;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750">Add to My Collection</a> &#124; <a class=remove" href=/correct;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750" title=Submit corrections for this document">Correct Errors</a></div> </div> <!-- End bottom information\_bar --> <div id=conclusion"> <div href=/viewdoc/similar;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750" title=Find documents related to this document">Related Documents</a>: <a href=/viewdoc/similar;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750&amp;type=ab" title=Documents that cite the same works">Active Bibliography</a> &#124; <a href=/viewdoc/similar;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750&amp;type=cc" title=Documents cited by the same works">Co-citation</a><!-- &#124; <a href=#">Version History</a> --></div> </div> <!-- End conclusion --> </div> <!-- End primary\_content - Opened in viewdoc\_header --> </div> <!-- End center\_content - Opened in viewdoc\_header --> <div id=footer"> <p><a href=/;jsessionid=5AC149882DBB363288302750A37F33E6" title=CiteSeerX Home Page">Home Page</a> &#124; <a href=/about/site;jsessionid=5AC149882DBB363288302750A37F33E6" title=Learn more about this scientific literature digital library and search engine">About CiteSeer</sup>x</sup></a> &#124; <a href=http://csxblog.ist.psu.edu/" title=Updates and New Features">Bulletin</a> &#124; <a href=/myciteseer/action/submitUrl;jsessionid=5AC149882DBB363288302750A37F33E6" title=Add

[Your Document](#) Submit Documents

[Feedback](#) [Privacy Policy](#) [CiteSeerX Data](#) [Get SeerSuite Source Code](#)

2007–2010 The Pennsylvania State University

Developed at and hosted by The College of Information Sciences and Technology at Penn State



CiteSeerX—Memory Performance of Master Go Players

gov.sandia.cortex.metadata.MetadataMap serialization=eustom unserializable-parents default loadFactor=0.75 threshold=12 int=16 int=5 string=Urls gov.sandia.cortex.metadata.Metadata id=99fd5419-208c-4350-8507-31b397f63f17 name=Urls data class=List

<http://citeseerx.ist.psu.edu/css/csxsecondary.css;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/favicon.ico;jsessionid=5AC149882DBB363288302750A37F33E6>

[http://citeseerx.ist.psu.edu/search\\_plugins/citeseerx\\_general.xml;jsessionid=5AC149882DBB363288302750A37F33E6](http://citeseerx.ist.psu.edu/search_plugins/citeseerx_general.xml;jsessionid=5AC149882DBB363288302750A37F33E6)

[http://citeseerx.ist.psu.edu/search\\_plugins/citeseerx\\_author.xml;jsessionid=5AC149882DBB363288302750A37F33E6](http://citeseerx.ist.psu.edu/search_plugins/citeseerx_author.xml;jsessionid=5AC149882DBB363288302750A37F33E6)

[http://citeseerx.ist.psu.edu/search\\_plugins/citeseerx\\_title.xml;jsessionid=5AC149882DBB363288302750A37F33E6](http://citeseerx.ist.psu.edu/search_plugins/citeseerx_title.xml;jsessionid=5AC149882DBB363288302750A37F33E6)

<http://citeseerx.ist.psu.edu/css/niftyCorners.css;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/js/niftycube.js;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/dwr/interface/MetadataCartJS.js;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/dwr/interface/GetContextJS.js;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/dwr/engine.js;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/js/metacart.js;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/js/mootools.js;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/js/mootabs.js;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/js/searchbox.js;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/js/citecontext.js;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/js/mousedout.js;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/css/wforms.css;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/js/correctionutils.js;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/>

<http://citeseerx.ist.psu.edu/about/site;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://csxblog.ist.psu.edu/>

<http://citeseerx.ist.psu.edu/submit;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/feedback;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/metacart;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/myciteseer/action/accountHome;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/images/blank.gif;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/search;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/search;jsessionid=5AC149882DBB363288302750A37F33E6?q=&sort=rel&submit=Search>

[http://citeseerx.ist.psu.edu/advanced\\_search;jsessionid=5AC149882DBB363288302750A37F33E6](http://citeseerx.ist.psu.edu/advanced_search;jsessionid=5AC149882DBB363288302750A37F33E6)

<http://citeseerx.ist.psu.edu/help/search;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/search;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/search;jsessionid=5AC149882DBB363288302750A37F33E6?q=&sort=cite&t=auth&submit=Search>

[http://citeseerx.ist.psu.edu/advanced\\_search;jsessionid=5AC149882DBB363288302750A37F33E6](http://citeseerx.ist.psu.edu/advanced_search;jsessionid=5AC149882DBB363288302750A37F33E6)

<http://citeseerx.ist.psu.edu/help/search;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/search;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/search;jsessionid=5AC149882DBB363288302750A37F33E6?q=&t=table&submit=Search>

<http://citeseerx.ist.psu.edu/help/search;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/viewdoc/summary;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750>

<http://citeseerx.ist.psu.edu/viewdoc/similar;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750>

<http://citeseerx.ist.psu.edu/viewdoc/similar;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750&type=ab>

<http://citeseerx.ist.psu.edu/viewdoc/similar;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750&type=cc>

<http://citeseerx.ist.psu.edu/viewdoc/versions;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750>

<http://citeseerx.ist.psu.edu/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750>

<http://www.itee.uq.edu.au/~janetw/ComputerGo/go-master.pdf>

<http://www.itee.uq.edu.au/~janetw/Computer%20Go/go-master.pdf>

<http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750&rep=rep1&type=pdf>

<http://citeseerx.ist.psu.edu/icons/pdf.gif;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/myciteseer/action/addPaperCollection;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750>

<http://citeseerx.ist.psu.edu/correct;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750>

<http://citeseerx.ist.psu.edu/myciteseer/action/editMonitors;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750>

<http://citeseerx.ist.psu.edu/myciteseer/action/editTags;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/myciteseer/action/editTags;jsessionid=5AC149882DBB363288302750A37F33E6?tag=&doi=10.1.1.130.9750&submit=Submit>

<http://citeseerx.ist.psu.edu/viewdoc/bookmark;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750&title=Memory Performance of Master Go Players&site=citeulike>

<http://citeseerx.ist.psu.edu/icons/bookmarks/citeulike.png;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/viewdoc/bookmark;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750&site=connotea>

<http://citeseerx.ist.psu.edu/icons/bookmarks/connotea.png;jsessionid=5AC149882DBB363288302750A37F33E6>

<http://citeseerx.ist.psu.edu/viewdoc/bookmark;jsessionid=5AC149882DBB363288302750A37F33E6>

?doi=10.1.1.130.9750&site=bibsonomy&title=Memory Performance of Master Go Players</url>  
 <url>http://citeseerx.ist.psu.edu/icons/bookmarks/bibsonomy.png;jsessionid=5AC149882DBB363288302750A37F33E6</url>  
 <url>http://citeseerx.ist.psu.edu/viewdoc/bookmark;jsessionid=5AC149882DBB363288302750A37F33E6  
 ?doi=10.1.1.130.9750&title=Memory Performance of Master Go Players&site=delicious</url>  
 <url>http://citeseerx.ist.psu.edu/icons/bookmarks/delicious.gif;jsessionid=5AC149882DBB363288302750A37F33E6</url>  
 <url>http://citeseerx.ist.psu.edu/viewdoc/bookmark;jsessionid=5AC149882DBB363288302750A37F33E6  
 ?doi=10.1.1.130.9750&title=Memory Performance of Master Go Players&site=digg</url>  
 <url>http://citeseerx.ist.psu.edu/icons/bookmarks/digg.png;jsessionid=5AC149882DBB363288302750A37F33E6</url>  
 <url>http://citeseerx.ist.psu.edu/viewdoc/bookmark;jsessionid=5AC149882DBB363288302750A37F33E6  
 ?doi=10.1.1.130.9750&title=Memory Performance of Master Go Players&site=reddit</url>  
 <url>http://citeseerx.ist.psu.edu/icons/bookmarks/reddit.gif;jsessionid=5AC149882DBB363288302750A37F33E6</url>  
 <url>http://citeseerx.ist.psu.edu/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=163788</url>  
 <url>http://citeseerx.ist.psu.edu/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=291193</url>  
 <url>http://citeseerx.ist.psu.edu/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=286400</url>  
 <url>http://citeseerx.ist.psu.edu/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=1320987</url>  
 <url>http://citeseerx.ist.psu.edu/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=1617061</url>  
 <url>http://citeseerx.ist.psu.edu/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=3455954</url>  
 <url>http://citeseerx.ist.psu.edu/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=1617060</url>  
 <url>http://citeseerx.ist.psu.edu/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=4964019</url>  
 <url>http://citeseerx.ist.psu.edu/viewdoc/summary;jsessionid=5AC149882DBB363288302750A37F33E6?cid=1957941</url>  
 <url>http://citeseerx.ist.psu.edu/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=5014568</url>  
 <url>http://citeseerx.ist.psu.edu/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=7209545</url>  
 <url>http://citeseerx.ist.psu.edu/showciting;jsessionid=5AC149882DBB363288302750A37F33E6?cid=7209546</url>  
 <url>http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750&rep=rep1&type=pdf</url>  
 <url>http://citeseerx.ist.psu.edu/myciteseer/action/addPaperCollection;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750</url>  
 <url>http://citeseerx.ist.psu.edu/correct;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750</url>  
 <url>http://citeseerx.ist.psu.edu/viewdoc/similar;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750</url>  
 <url>http://citeseerx.ist.psu.edu/viewdoc/similar;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750&type=ab</url>  
 <url>http://citeseerx.ist.psu.edu/viewdoc/similar;jsessionid=5AC149882DBB363288302750A37F33E6?doi=10.1.1.130.9750&type=cc</url>  
 <url>http://citeseerx.ist.psu.edu/;jsessionid=5AC149882DBB363288302750A37F33E6</url>  
 <url>http://citeseerx.ist.psu.edu/about/site;jsessionid=5AC149882DBB363288302750A37F33E6</url>  
 <url>http://csxblog.ist.psu.edu/</url>  
 <url>http://citeseerx.ist.psu.edu/myciteseer/action/submitUrl;jsessionid=5AC149882DBB363288302750A

37F33E6</url>  
 <url>http://citeseerx.ist.psu.edu/feedback;jsessionid=5AC149882DBB363288302750A37F33E6</url>  
 <url>http://citeseerx.ist.psu.edu/privacy;jsessionid=5AC149882DBB363288302750A37F33E6</url>  
 <url>http://citeseerx.ist.psu.edu/about/metadata;jsessionid=5AC149882DBB363288302750A37F33E6</url>  
 > <url>http://sourceforge.net/projects/citeseerx/</url> <url>http://www.psu.edu</url>  
 <url>http://ist.psu.edu/</url> <url>http://ist.psu.edu/</url>  
 <url>http://citeseerx.ist.psu.edu/images/istlogo.gif;jsessionid=5AC149882DBB363288302750A37F33E6</url>  
 </data> </gov.sandia.cortex.metadata.Metadata> <string>Content-Encoding</string> -  
 <gov.sandia.cortex.metadata.ReadableMetadata> <id>7917ffe9-608e-49e5-9b35-2cccd1dfa63aContent-Encoding</id> <name>Content-Encoding</name> <data class=-string>ISO-8859-1</data>  
 <readableData>ISO-8859-1</readableData> </gov.sandia.cortex.metadata.ReadableMetadata>  
 <string>FilePath</string> -<gov.sandia.cortex.metadata.ReadableMetadata> <id>9547160b-4528-435c-b78d-0f6cfb9732e0</id> <name>FilePath</name> <data class=-string>http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.130.9750</data>  
 <readableData>http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.130.9750</readableData>  
 </gov.sandia.cortex.metadata.ReadableMetadata> <string>Similarity</string> -  
 <gov.sandia.cortex.metadata.ReadableMetadata> <id>390509aa-50a1-4fed-97c7-ce66eb6b9d2b</id>  
 <name>Similarity</name> <data class=-double>0.046708903646188765</data>  
 <readableData>0.046708903646188765</readableData>  
 </gov.sandia.cortex.metadata.ReadableMetadata> <string>Content-Type</string> -  
 <gov.sandia.cortex.metadata.ReadableMetadata> <id>7917ffe9-608e-49e5-9b35-2cccd1dfa63aContent-Type</id> <name>Content-Type</name> <data class=-string>text/plain</data>  
 <readableData>text/plain</readableData> </gov.sandia.cortex.metadata.ReadableMetadata> </map> -  
 <gov.sandia.cortex.metadata.MetadataMap> -<default> <targetID>7917ffe9-608e-49e5-9b35-2cccd1dfa63a</targetID> </default> </gov.sandia.cortex.metadata.MetadataMap>  
 </gov.sandia.cortex.metadata.MetadataMap> </gov.sandia.cortex.document.Document>

## APPENDIX E. ANALYSIS ON ORIGINAL STANLEY RESULTS

This appendix contains an analysis of the original STANLEY results. A major disparity, and one that could dramatically affect the comparison, was the number of scores for each site that were returned as zero (0.000). This could be attributed to the tool being unable to score the document (i.e., it is a picture) or there are no words in common except for the common stop words such as “and” or “the.” It could also be due to weaknesses in the classifiers, but this would generally not lead to scores of zero. Regardless of the cause, the number of zero scores was significant in some cases. For example, Bloomberg.com returned 43 scores for March, 2010. Of these 43 scores, 20 were zero. Likewise, for May, [www.nyt.com](http://www.nyt.com) returned 26 sites; 24 were zero. Similar results were encountered across all the websites returned. This thesis assumed that these scores were due to tool’s inability to score the document, meaning that the site was text free. The reader will notice that the correlation and fit values presented here are lower than those in section IV.C.

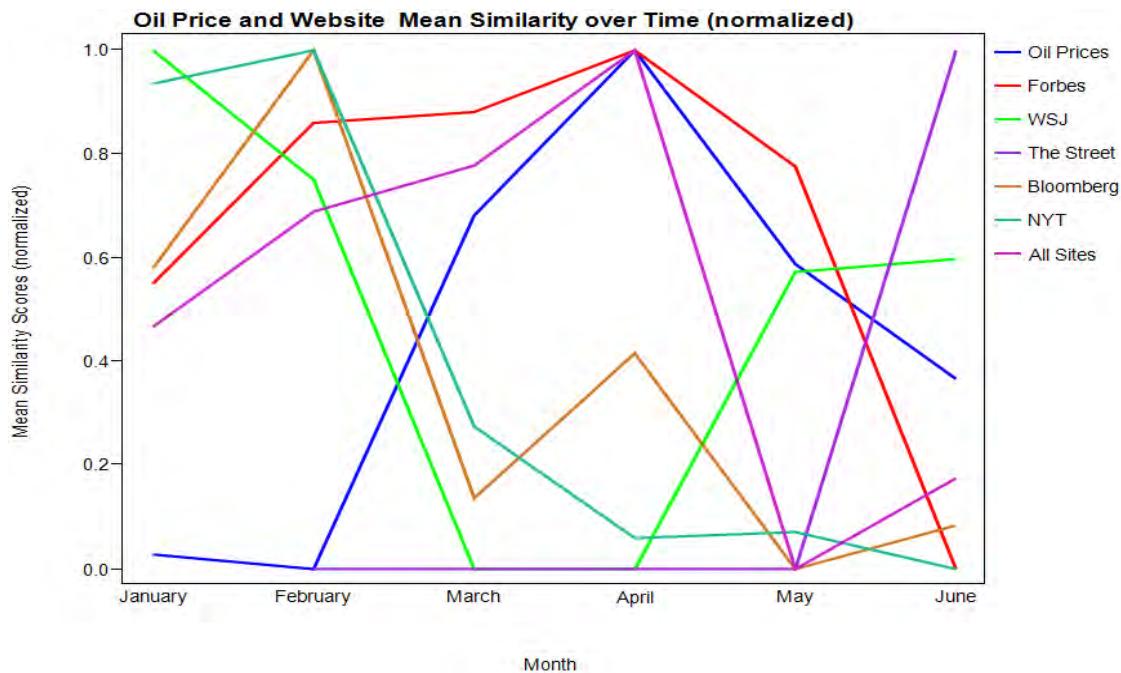


Figure 52. Mean and median scores with oil prices over time (normalized) (best viewed in color)

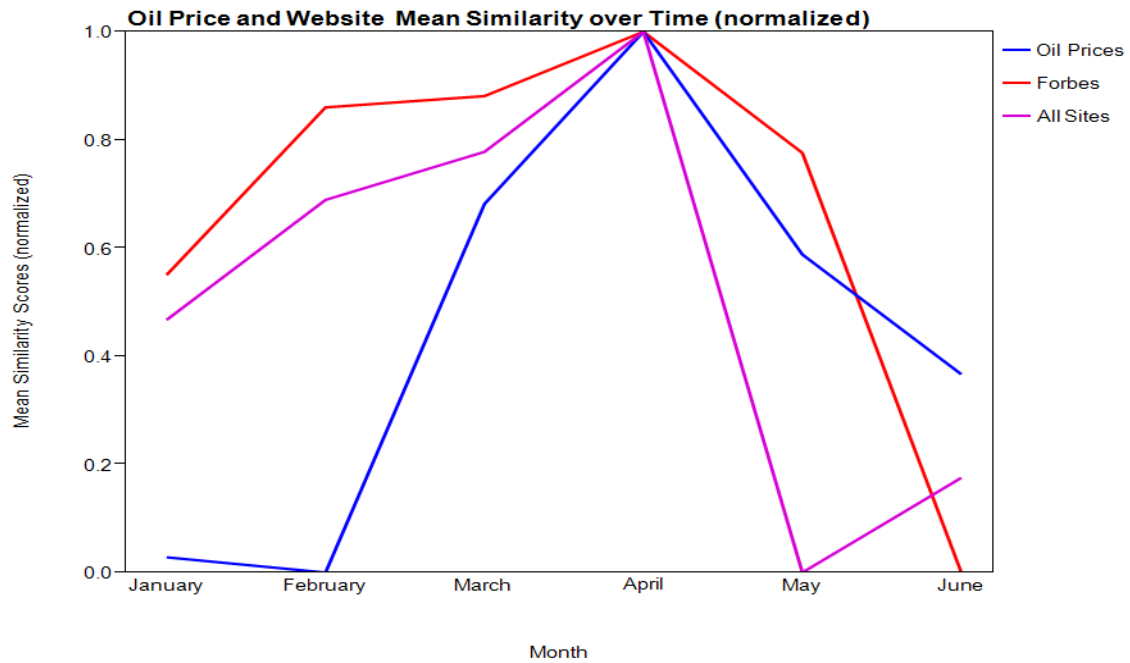


Figure 53. Oil prices, forbes.com and all sites score (means, normalized) (best viewed in color)

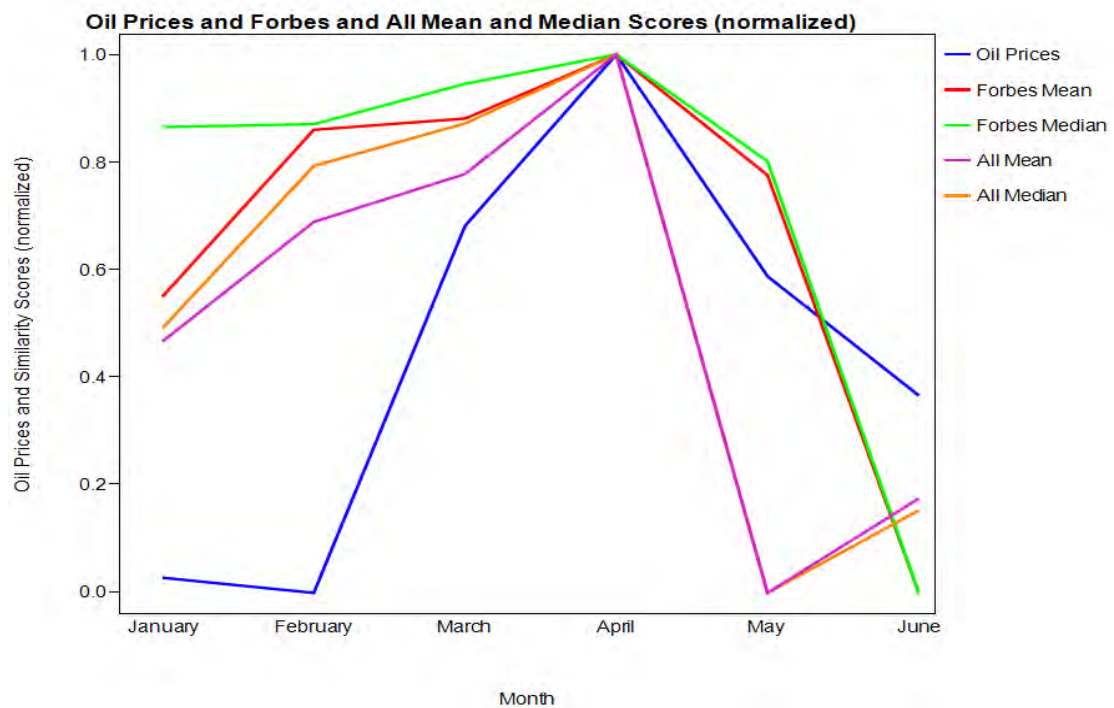


Figure 54. Mean and median scores with oil prices over time (normalized) (best viewed in color)

Correlations					
	Oil Prices Normalized	ForbesMeanNormalized	ForbesMedianNormalized	AllMeanNormalized	AllMedianNormalized
Oil Prices Normalized	1.0000	0.3741	0.2054	0.3067	0.2447
ForbesMeanNormalized	0.3741	1.0000	0.9447	0.6192	0.6433
ForbesMedianNormalized	0.2054	0.9447	1.0000	0.5902	0.6163
AllMeanNormalized	0.3067	0.6192	0.5902	1.0000	0.9935
AllMedianNormalized	0.2447	0.6433	0.6163	0.9935	1.0000

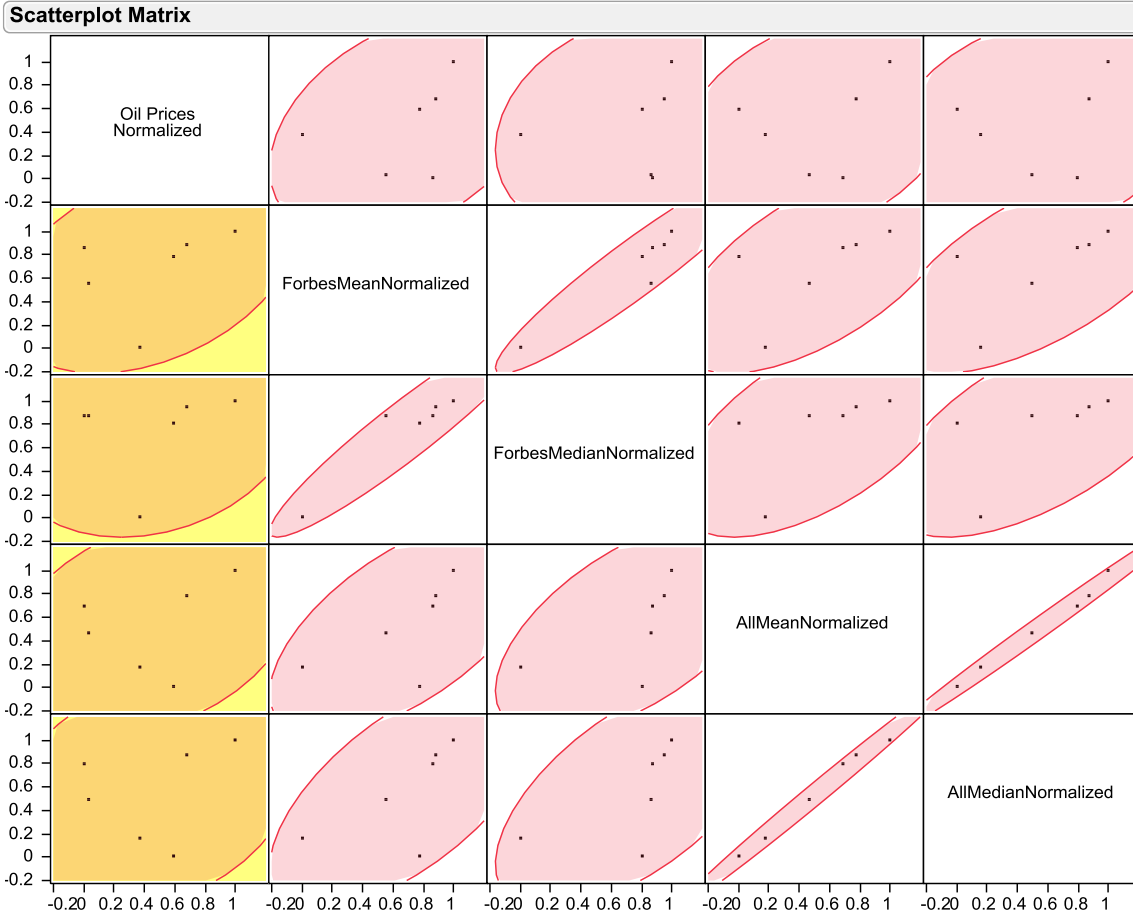


Figure 55. Scatterplot matrix showing correlations of oil price to similarity scores

Pairwise Correlations						
Variable	by Variable	Correlation	Count	Lower 95%	Upper 95%	Signif Prob
ForbesMeanNormalized	Oil Prices Normalized	0.3741	6	-0.6282	0.9095	0.4651
ForbesMedianNormalized	Oil Prices Normalized	0.2054	6	-0.7274	0.8717	0.6962
ForbesMedianNormalized	ForbesMeanNormalized	0.9447	6	0.5705	0.9941	0.0045 *
AllMeanNormalized	Oil Prices Normalized	0.3067	6	-0.6722	0.8954	0.5544
AllMeanNormalized	ForbesMeanNormalized	0.6192	6	-0.3867	0.9522	0.1899
AllMeanNormalized	ForbesMedianNormalized	0.5902	6	-0.4248	0.9478	0.2175
AllMedianNormalized	Oil Prices Normalized	0.2447	6	-0.7073	0.8813	0.6402
AllMedianNormalized	ForbesMeanNormalized	0.6433	6	-0.3520	0.9558	0.1681
AllMedianNormalized	ForbesMedianNormalized	0.6163	6	-0.3907	0.9518	0.1926
AllMedianNormalized	AllMeanNormalized	0.9935	6	0.9392	0.9993	<.0001 *

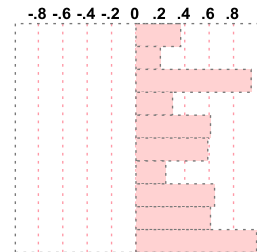


Figure 56. Pair-wise correlations among oil prices and similarity scores

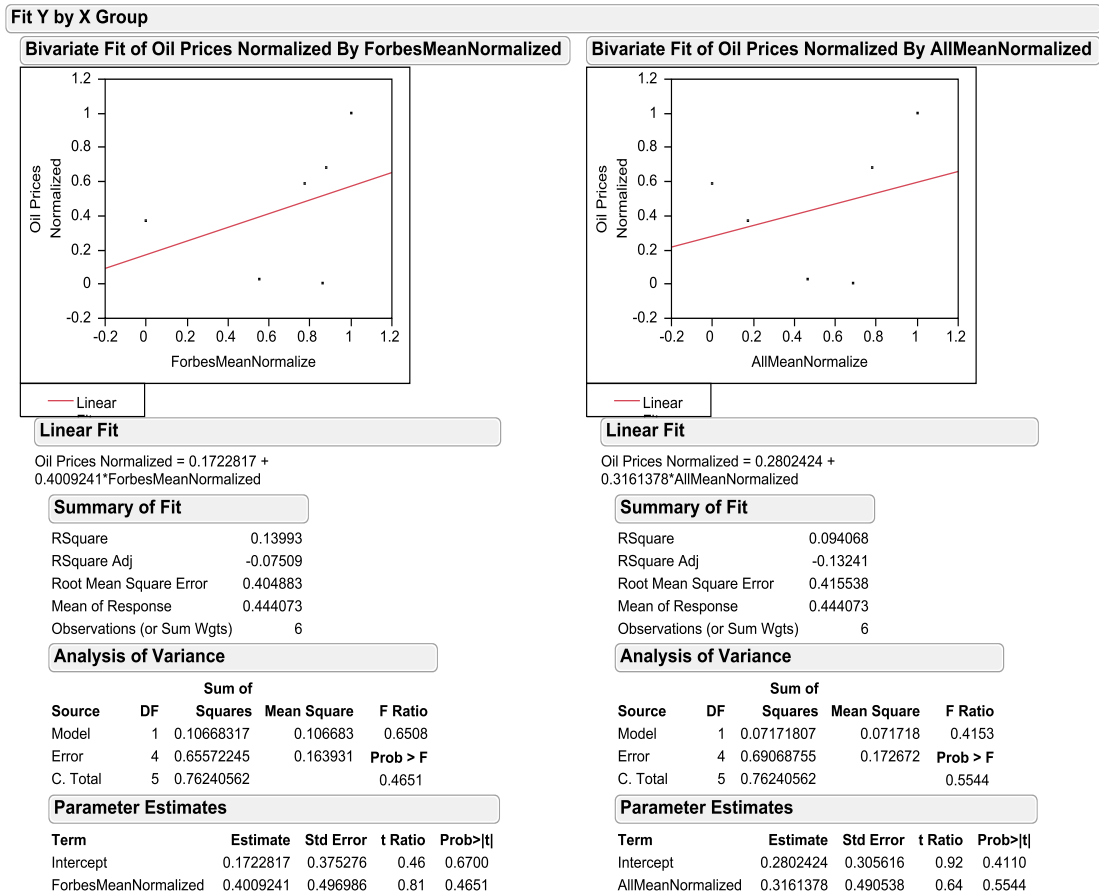


Figure 57. Regression analysis for oil prices with mean similarity scores (Forbes and All)

Response Oil Prices Normalized					
Summary of Fit					
RSquare		0.149071			
RSquare Adj		-0.41821			
Root Mean Square Error		0.465028			
Mean of Response		0.444073			
Observations (or Sum Wgts)		6			
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	
Model	2	0.11365288	0.056826	0.2628	
Error	3	0.64875274	0.216251		Prob > F
C. Total	5	0.76240562		0.7849	
Parameter Estimates					
Term		Estimate	Std Error	t Ratio	Prob> t
Intercept		0.1620215	0.434795	0.37	0.7342
ForbesMeanNormalized		0.3201147	0.726938	0.44	0.6895
AllMeanNormalized		0.1255087	0.69911	0.18	0.8690
Effect Tests					
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
ForbesMeanNormalized	1	1	0.04193481	0.1939	0.6895
AllMeanNormalized	1	1	0.00696972	0.0322	0.8690

Figure 58. Summary of fit results for oil prices and Forbes and All mean similarity scores (combined)

When the zero scores were removed from the results, there was a significant increase in correlation and R-squared values between the two sets of data. For the [www.forbes.com](http://www.forbes.com) mean similarity regression, the R-squared value increased from 0.140 to 0.415, an increase of nearly 300%. This is a significant increase and is intriguing and deserves some further investigation in future research.

THIS PAGE INTENTIONALLY LEFT BLANK

## APPENDIX F. SURVEY DATA DEVELOPMENT TOOL USERS GUIDE

To generate the case files for the survey data:

- 1) Determine if you are going to use previously run survey settings or new settings. If previously used survey settings are to be used, the surveydatacasefilegenerator.ini MUST contain the desired session's settings. See section 12) below for information regarding the saving of settings and the .ini file. If new settings are to be run, ensure that the surveydatacasefilegenerator.ini file is deleted from the directory, if one exists.
- 2) Double click the SurveyCaseFileGenerator.bat file to launch the generator tool.
- 3) A "Survey Data Tool - Case file Generator" window will pop up.
- 4) If a previously run setting is being used, and you simply want to modify the parameters, go to step 9).
- 5) The first field prompts you to specify the directory where Survey Data is located. Click on the file browse button next to the field, a "Load Survey Data" file chooser window will pop up. Navigate to the appropriate root directory where all the survey data files are located. The input data files should be in the \*.csv format. They should all be under the same root directory. You can also select a single data file. In that case, the tool would only load the data from that particular data file.
- 6) Click the "Load" button from the file chooser window after the directory or file is selected. This will dismiss the file chooser window and cause the tool to update the "Term Types" table. By default, all terms are type "Categorical." To change a term to "Continuous," click on the appropriate "Type" cell, and select the desired type from an option menu. You can also specify a particular term as a "Target" by selecting the appropriate "Target" cell. If a term is selected as the "Target," it will not be ranked since its values are used to calculate Two-Way Classification Chi-Square for other "Categorical" terms.

- 7) If you DO NOT specify a “Target” term, you can customize the distributions for Categorical terms. To do that, click on the “Customize Target Distributions for Categorical Terms” button below the “Term Types” table. This customizes the expected distribution for the terms. By default, all categorical terms are evenly distributed, i.e., if a categorical term has values either 1 or 2, by default, the tool expects 50% of the term is valued as 1, and 50% of the term is valued as 2. You can alter the distribution to match expected output for each term.
- 8) After you specify/customize the characteristics of the terms, click the “Load” button underneath the file or directory selection space towards the top of the interface. The tool will read through all the data files you specified here, whether a single file or all the \*.csv files located under the specified directory. Depending on the size of the data, this may take a while. But when all the data is read and loaded, the tool will update the “Terms Ranking Table.” For Categorical terms, the ranking is determined by comparing the chi-square values, and for Continuous terms, the ranking is determined by comparing the variance. By default, only the top 4 ranked terms are selected. You can select all the terms, or deselect a term or select another term by selecting the appropriate “chi-square/variance” cell. The selected terms will affect what terms are used in stereotype generation and which are used for data importation.
- 9) See Figure 59

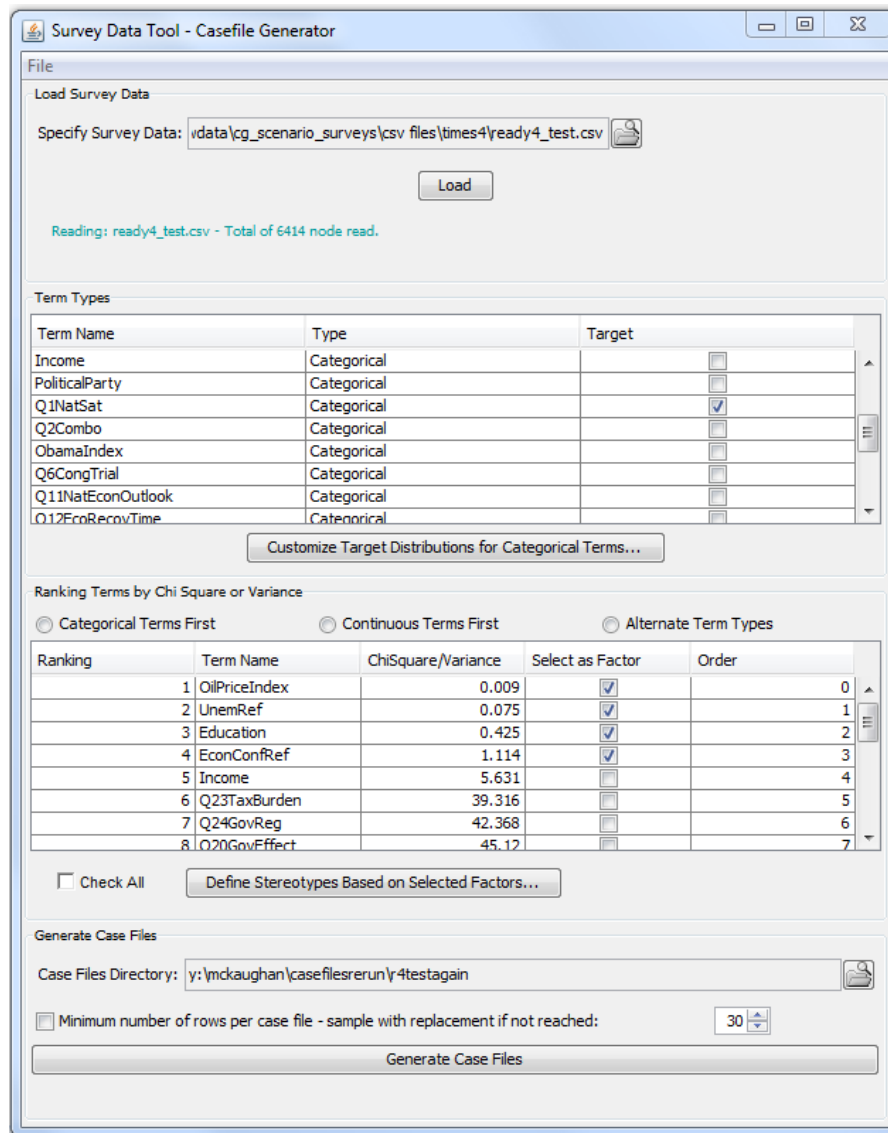


Figure 59. Survey tool with ability to order terms

- 10) After you specify the exact terms you want to included in the construction of stereotypes, it is a good idea to customize the definition of the “bins” for each selected terms. To do that, you should click the “Define Stereotypes Based on Selected Factors...” button below the Terms Ranking Table. This will bring up another dialog window that lists all the terms you’ve selected. By default, an individual value (survey response) from a categorical term is set as a “bin,” i.e., if the term “Race” has 3 values: 1, 2, and 3, three “Bins” should be defined for that term. The tool initially assumes all terms contain 6 bins. This is adjustable. To

change, deselect the “Use each value to designate a demographic” bin toggle button. Note that all checked terms from 6) above will be displayed in the new window. Only terms that are desired for stereotype generation should be changed in this step, unless you desire to alter the format or name of the survey responses as they will appear in the case files. For terms that will be included in the stereotype, check the “Stereotype” button. Then, edit the table to customize the values for each “Bin.” To add another row or delete an old one, you should change the value in the “Number of Demographic Bins” numerical spinner. For each “Bin,” the “Short Name” blank will determine the naming convention for that portion of the stereotype, so choose a clear, unambiguous short name.

- 11) Once you customized all the bins, click the “Apply” button on the dialog window. Once all the demographic bins are defined, the “Generate Case File” button on the Survey Data Tool will become enabled. Before generating the case files, specify a Case Files Directory where all the generated case files should reside. Ensure the directory where the files will be saved is new or empty, as case files will not be overwritten. See Figure 60.

**OilPriceIndex Category**

☒ Use each value to designate a demographic bin ☐ Stereotype

Number of Demographic Bins: 6

Long Name	Short Name	Range (Low)	Range (High)
OilPriceIndexPo...	OIPoS Trend	PosTrend	PosTrend
OilPriceIndexNe...	OINegTrend	NegTrend	NegTrend

**UnemRef Category**

☒ Use each value to designate a demographic bin ☐ Stereotype

Number of Demographic Bins: 6

Long Name	Short Name	Range (Low)	Range (High)
UnemRefPosTrend	UNPosTrend	PosTrend	PosTrend
UnemRefNegTrend	UNNegTrend	NegTrend	NegTrend

**Education Category**

☐ Use each value to designate a demographic bin ☒ Stereotype

Number of Demographic Bins: 4

Long Name	Short Name	Range (Low)	Range (High)
EducationHighSc...	EDHighSchool	HighSchool	HighSchool
EducationSome...	EDSomeCollege	SomeCollege	SomeCollege
EducationGTColl...	EDGTCollge	GTCollge	GTCollge
EducationLTHigh...	EDLTHighSchool	LTHighSchool	LTHighSchool

**EconConfRef Category**

☒ Use each value to designate a demographic bin ☐ Stereotype

Number of Demographic Bins: 6

Long Name	Short Name	Range (Low)	Range (High)
EconConfRefNe...	ECNegTrend	NegTrend	NegTrend
EconConfRefPo...	ECPoS Trend	PosTrend	PosTrend

Apply Cancel

Figure 60. Bin delineation GUI

- 12) You can also specify a minimum number of rows per case file. If selected, unless a case file for the stereotype has empty rows to begin with, the tool will perform sample with replacement on the existing rows to generate up to minimum number of rows for that case file.
- 13) At any time in the above process, but most probable once the bins have been delineated, you can save the settings. This is done by exiting the tool by pressing the red **—X** at the upper right of the tool's GUI. You have to exit this way for the settings to be saved. After a few moments, while the tool is updating the .ini file,

the tool will exit. Upon reopening the tool, the last settings will be present. To run the tool with those settings, press the “Define Stereotypes Based on Selected Factors” button, press “Apply” in that GUI, then rerun the tool by pressing the “Generate Case Files” button.

NOTE: It is advisable to save your settings under recognizable names. Make a copy of the surveydatatoolcasefilegenerator.ini file. Rename it as something you will recognize. If you desire to use these settings again, you MUST rename the file as surveydatatoolcasefilegenerator.ini, open the tool, open the “Define Stereotypes...” GUI, and press “Apply.” Then you can generate case files.

## LIST OF REFERENCES

- 2010 current events. (2010). Retrieved May 13, 2011, from <http://www.infoplease.com/news/2010/current-events/>
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Processes*, 50, 179–180–211. Doi: 0749-5978/91
- Alt, J. K., Everton, S., & Lieberman, S. (2009). *Violent extremist network representation and attach the network course of action analysis in social simulation*. Monterey, CA: Modeling, Virtual Environment Simulations (MOVES) Institute, Naval Postgraduate School.
- Alt, J. K., Jackson, L. A., Hudak, D., & Lieberman, S. (2009). The cultural geography model: Evaluating the impact of tactical operational outcomes on a civilian population in an irregular warfare environment. *Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, 6(4), 185–199. Doi: 10.1177/1548512909355000
- Alt, J. K., & Lieberman, S. (2010). Developing cognitive models for social simulation from survey data. In Chai, S., Salerno, J., & Mabry, P. (eds.) *Advances in Social Computing, Lecture Notes in Computer Science* (pp. 323–329). Heidelberg: Springer Berlin.
- Alt, J. K., & Lieberman, S. (2010). Modeling the theory of planned behavior from survey data for action choice in social simulations. In the *Proceedings of the 19th Conference on Behavior Representation in Modeling and Simulation*, Charleston, SC, 126–133, March 2010.
- Alt, J. K., Lieberman, S., McClain, J. T., & Heath, Z. (2010). *Leveraging a document corpus for human terrain system modeling*. [Unpublished manuscript].
- Baez, F. (2011). *Data transformation and tools for the Cultural (CG) Model*. [PowerPoint slides].
- Baeza-Yates, R., & Ribeiro-neto, B. (1999). *Modern information retrieval*. Reading, MA: Addison-Wesley.
- Basilico, J. D., Diegert, C. F., Heath, Z., Ko, T. H., Linebarger, J. M., Pancerella, C. M., Schneider, P. A. W. (2008). *Large scale social simulation*. SAND 2008–7831). Albuquerque, New Mexico: Sandia National Laboratories.

- Bauer, T., Laham, D., Benz, Z., Dooley, S., Kimmel, J., & Oberbrekli, R. (2005). Text analysis and dimensionality reduction for automatically populating a cognitive model framework. In Forsythe, M. L. Bernard and T. E. Goldsmith (eds.) *Cognitive Systems : Human Cognitive Models in Systems Design* (pp. 153–176). Mahwah, NJ: Lawrence Erlbaum Associates.
- Berry, N., Ko, T. H., Lee, M., Moy, T., Pickett, M., Smrcka, J., Wu, B., et al. (2003). *Computational social dynamic modeling of group recruitment*. SAND2003–8754. Albuquerque, NM: Sandia National Laboratories.
- Bug Man. *Economic confidence index*. Retrieved April 2, 2011, from <http://www.quantshare.com/item-802-economic-confidence-index>
- Complex adaptive systems*. Retrieved November 15, 2010, from [http://en.wikipedia.org/wiki/Complex\\_adaptive\\_system](http://en.wikipedia.org/wiki/Complex_adaptive_system)
- Department of Defense (DoD). (2007). *Joint Operating Concept (JOC), Irregular warfare (IW)*. Washington D.C.
- Dumais, S. (1991). Improving the retrieval of information from external sources. *Instruments and Computers*, 23(2), 229–236.
- Effective data mining using JMP partition platform*. Retrieved July 19, 2011, from [http://www.jmp.com/software/whitepapers/partition\\_platform/index.shtml](http://www.jmp.com/software/whitepapers/partition_platform/index.shtml)
- Fischer, W. (1989). Clarifying the narrative paradigm. *Communication Monographs*, 56, 55–58.
- Gutenberg Project*. (2010a). Retrieved October 25, 2010, from [http://www.gutenberg.org/wiki/Main\\_Page](http://www.gutenberg.org/wiki/Main_Page)
- Gutenberg Project*. (2010b). Retrieved October 25, 2010, from [http://en.wikipedia.org/wiki/Project\\_Gutenberg](http://en.wikipedia.org/wiki/Project_Gutenberg)
- Hudak, D., Baez, F., Jones, S., Perkins, T. K., Duong, D., Willis, J., Pearman, G., et al. (2010). *Cultural modeling support to Pakistan-Afghanistan (PAKAF) strategic multi-layered assessment*. Monterey, CA: U.S. Army Training and Doctrine Command.
- Jackson, L. A., & Perkins, T. K. (2009). *Team 12: Cultural geography modeling and analysis for IDFW 18*.
- JMP*. (2010). Retrieved August 10, 2011, from <http://www.jmp.com/index.shtml>
- Kalton, G. (1983). *Introduction to survey sampling*. Beverly Hills: Sage.

- Kielcolt, K., & Nathan, L. (1985). *Secondary analysis of survey data*. Beverly Hills: Sage.
- Kilkullen, D. (2007). *Counterinsurgency in Iraq: Theory and practice*. Retrieved May 24, 2010, from <http://smallwarsjournal.com/documents/kilcullen1.pdf>
- Lieberman, S., & Alt, J. K. (2010). Developing social networks for artificial societies from survey data. In Chai, S., Salerno, J., & Mabry, P. (eds.) *Advances in Social Computing, Lecture Notes in Computer Science* (pp. 159–168). Heidelberg: Springer Berlin.
- Mansoor, P. R., & Ulrich, M. S. (2007). Linking doctrine to action: A new COIN center of gravity analysis. *Military Review*, (September-October), 45–46–51.
- McKaughan, D., Alt, J. K., Heath, Z., & McClain, J. (2011). Using a text analysis and categorization tool to generate Bayesian belief networks for use in cognitive social simulation from a document corpus. In *Proceedings of the Spring Simulation Multiconference*, Boston MA, April 2011.
- National Research Council. (2008). *Behavioral modeling and simulation: From individuals to societies* National Academies Press.
- Netica application*. (2011). Retrieved June 3, 2011, from <http://www.norsys.com/netica.html>
- Osgood, J. *A study of Clausewitz's concept of the military center of gravity*. Retrieved July 23, 2011, from <http://www.juris99.com/mil/wc4.htm>
- Pant, G., Srinivasan, P., Menczer, F. (2011). *Crawling the web*. Retrieved from <http://dollar.biz.uiowa.edu/~pant/Papers/crawling.pdf>
- Ross, A. *Middle East and North African turmoil raised the oil prices*. Retrieved from <http://topnews.us/content/238367-middle-east-and-north-african-turmoil-raised-oil-prices>
- Salton, G., & Buckley, C. (1991). Automatic text structuring and retrieval experiments in automatic encyclopedia searching. In *Proceeding SIGIR '91 Proceedings of the 14<sup>th</sup> Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.
- Salton, G., Wong, A., & Yang, C. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 613–614–620.
- Silverman, B. G. (2007). *Human terrain data: What should we do with it?*

- Sun, R. (2007). *Cognitive social simulation incorporating cognitive architectures*. Retrieved 2/5, 2011, from <http://www.cogsci.rpi.edu/~rsun/folder-files/sun-IS-2007.pdf>
- Weka*. (2010). Retrieved October 24, 2010, from <http://www.cs.waikato.ac.nz/~ml/weka/>
- Wiedemann, M. (2010). *Robust parameter design for agent-based simulation models with application in a cultural geography model*. (Unpublished MS). Naval Post Graduate School, Monterey, CA.
- World values survey*. Retrieved December 8, 2010, from <http://www.worldvaluessurvey.org/>

## **INITIAL DISTRIBUTION LIST**

1. Defense Technical Information Center  
Ft. Belvoir, Virginia
2. Dudley Knox Library  
Naval Postgraduate School  
Monterey, California
3. LTC Jonathan Alt, USA, Director  
U.S Army TRADOC Analysis Center  
Monterey, California
4. CPT Richard Brown, USA  
U.S. Army TRADOC Analysis Center  
Monterey, California